THE CLONING AND CHARACTERIZATION OF MICROSATELLITE POLYMORPHISMS FROM CHROMOSOME 21

by

Vikram Sharma

A DISSERTATION

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

Doctor of Philosophy

September 1992

APPROVED:

[redacted]

(Professor in Charge of Thesis)

[redacted]

(Chairman, Graduate Council)

To

Gun, Ved and Krishna

ACKNOWLEDGEMENTS

I wish to thank Mike Litt, David Browne, Jeff Luty, Marilyn Jones, Gillian Harron, Keith Wilson, Tom Allen, Lisa Parshley, Haydeh Payami, Ruth Litt, Xue Ya Hauge, Bob Koler, Thas Promchotikul, Rachel Tsu and all my friends in the Zonana and Magenis labs for their unflagging support and friendship towards me.

TABLE OF CONTENTS

FIGURES

TABLES

ABSTRACT

This dissertation addresses itself to finding and characterizing polymorphic DNA markers. These are essential for constructing genetic linkage maps, by which inherited disease genes become localized to a chromosome region- an important first step if a gene is to be cloned. Genetic maps are based on observing recombination between polymorphic loci as a result of meiosis. The frequency with which loci recombine is a measure of their distance. Computerized likelihood analysis of genotypes produced in multigeneration (CEPH) families determines map position and distance between loci by demonstrating linkage.

Currently many gaps exist in linkage maps reducing their effectiveness in mapping disease genes. New markers will benefit international and national collaborative efforts to construct and improve genetic linkage maps of the human genome. A number of important genetic conditions including familial Alzheimer's disease (FAD), Down's syndrome (DS) and amyotrophic lateral sclerosis (ALS), provide the impetus to develop genetic markers from chromosome 21.

Microsatellites are a recently discovered class of DNA length polymorphisms from which genotypes are easily determined by polymerase chain reaction (PCR). We have demonstrated that the abundance, distribution and informativeness of (CA)n dinucleotide repeats makes them extremely valuable for genetic map constructions. To clone polymorphic sequences from chromsome 21, genomic libraries consisting of flow-sorted chromosome 21 DNA cloned into cosmids were screened for the presence of short tandem repeats (STR). Fifteen polymorphic sequences were isolated, characterized and mapped to human chromosomes; ten markers have heterozygosities greater than 0.70. Eight microsatellites are from chromosome 21; 4 map to the proximal half of the long arm of chromosome 21- an area which is poor in number of genetic markers and implicated in

familial Alzheimer's disease. Seven repeats suitable for linkage studies were used to construct a genetic linkage map of chromosome 21. This mapping data was incorporated into linkage maps currently being constructed for the US Human Genome Initiative and by the Center for the Study of Human Polymorphisms (CEPH). Seven different microsatellites isolated from the chromosome 21 flow sorted libraries mapped to other human chromosomes. A majority of these are highly informative, suitable for linkage studies and have been incorporated into the genetic map of their respective chromosome.

In the course of cloning and characterizing the polymorphisms described in this study, new techniques were developed which facilitate the ease with which microsatellite polymorphisms can be characterized and used to determine the genotypes of large numbers of people which are often required for linkage studies. These are primarily the use of formamide DNA sequencing gels, the transfering of amplified alleles to membranes and visualizing them with repeat or locus specific oligonucleotide probes. Increased throughput of samples is achieved by the use of rapid thermal cyclers adapted to 96 well microtiter plate format , loading sequencing gels multiple times and the use of multichannel pipetting devices.

I: INTRODUCTION

A: Genetic Linkage

Linkage Maps & Positional Cloning

The genes resonsible for a number of human genetic conditions (diseases) have been cloned. Virtually all of these were identified on the basis of pre-existing information about the biochemical basis of the disorder. Denoted "functional" cloning, "several hundred disease genes have been isolated since the advent of recombinant DNA technology 15 years ago" (Collins, 1992). While these are triumphs of modern science, more than 5000 human genetic syndromes have been described (McKusick, 1990). The vast majority of these do not provide biochemical clues and have not been localized to chromosomes, cloned or characterized with resolution sufficient to describe the underlying molecular defect of the disease. Nonetheless, modern techniques have a tremendous potential for reducing the impact of major human diseases on individuals, on families and on society. This will undoubtedly remain the central focus of molecular genetics in the coming decades.

Genetic linkage groups are blocks of genes which segregate together. They were first recognized as traits which segregated together. Early geneticists understood the basis for linkage groups when chromosomes were discovered (Smith, 1986). Traditionally, organisms (pea, fly, mouse) have been used to study genetics by controlled, experimental breeding. This is not done with humans. Although we recognize many mutant human phenotypes, the genetic information must be extracted from families as they are found in the population. There have been three methods traditionally used for mapping loci: the family method (linkage), somatic cell hybridization and cytologic characterization of chromosome rearrangements. Autosomal linkage groups have usually been defined by loci which consist of diseases whose genotype specifies a clinical condition; expressed genes which give rise to antigens, enzymes and protein isoforms; and chromosome structural rearrangements and heteromorphisms--most of which requires elaborate biochemical, cytologic or immunological methods for their detection (Smith, 1986; Hudson *et al.*, 1992).

As more loci were discovered genetic "maps" began to emerge which measured the distance between loci cytologically or by the degree of crossing over between loci segregating in offspring (recombination). Until recently, the development of better characterized genetic linkage maps has been limited by the availability of genetic markers (White *et al.*, 1985). Today, most markers are cloned, polymorphic segments of DNA with known or unknown function.

A decade has passed since Botstein *et al.* published their seminal work describing the use of restriction fragment length polymorphisms (RFLPs) to construct genetic linkage maps in humans (Botstein *et al.* 1980; Litt, 1991). The fundamental purpose of such maps is to find polymorphic markers which are linked to a trait or human genetic disease. This is done by demonstrating concordant segregation between a marker and a disease in families. Since markers on genetic maps ultimately correspond to physical positions on human chromosomes, a disease gene mapped by linkage analysis becomes amenable to characterization by molecular biological approaches. By constructing fine scale linkage and physical maps, genes can be mapped to small intervals of DNA and eventually cloned. Unlike functional cloning, the mapping of disease genes by linkage analysis does not require that the biochemical or genetic nature of the defect be known, nor that the polymorphic marker derive from part of the gene (Botstein *et al.*, 1980). This revolutionary approach to elucidating the molecular basis for inherited human diseases is called positional cloning. The success of positional cloning has been demonstrated by the cloning of the genes responsible for cystic fibrosis, neurofibromatosis I and myotonic dystrophy (Collins, 1992). In addition to serving as landmarks for positional cloning and physical mapping studies, markers linked to known genetic disorders can be used to determine the genotypes of individuals at risk. Preclinical detection of a genetic disorder may benefit treatment, allow prenatal testing and family counselling for a disorder whose impact may be predictable. Markers may help delineate between genetic and nongenetic components to complex disorders which appear to be inherited in subsets of families within the population.

To facilitate mapping disease genes by positional cloning a number of national and international collaborative research groups have begun to construct genetic linkage maps. The Center for the Study of Human Polymorphisms (CEPH) in Paris has collected 40 large multigeneration families and made them available to researchers developing genetic markers (Dausset *et al.*,1990). The genotypic data is returned to CEPH which uses it to construct linkage maps which are made available to participants. The US genome project is an ambitious program to completely sequence the $3.3x10^9$ nts. of DNA which comprise the haploid human genome. Highly dependent on technological innovations in DNA sequencing, data managment, physical and genetic map making, it is felt that this can be achieved within 20 years (Stephens *et al.*, 1990).

One of the first goals of the US human genome project is to construct a framework linkage map, consisting of markers with minimum heterozygosities of 70%. These highly informative (index) markers will be spaced along all human chromosomes, with distances between markers not to exceed 15 centimorgans (cM). A map of this resolution would require ≈300 markers. Index markers will be based on efficient, new protocols for determining the genotypes of individuals (i.e. PCR), so that linkage analysis with the full complement of markers will be more practical and efficient. A framework map will also allow additional markers to be positioned easily and accurately. As more markers are developed, the resolution of the linkage map will improve. Eventually a high resolution map will emerge with gaps between markers that do not exceed 2 cM.

Genetic distances between two loci are expressed in units of Morgans. This is defined as the number of expected crossovers occuring on a single chromatid between loci (Smith, 1986; Conneally & Rivas, 1980). Calculations of the average number of crossovers (visualized as chiasma) per haploid human genome yield an autosomal map length of 33 Morgans (or 3300cM) when averaged between the sexes. In small intervals where multiple crossovers will be very rare, crossovers can be detected as "recombinants". The genetic distance between adjacent loci is their recombination frequency (theta). Theta

equals the frequency of recombinant chromosomes out of the total number of recombinant and non recombinant chromosomes observed: theta= R ÷(NR+R). A recombination frequency of10% between two close loci is equivalent to a map distance of 0.10 morgans or 10 cM. A recombination frequency of 0.5 is observed between unlinked loci. When considering larger map distances, recombination frequencies become non linear and non additive due to the presence of multiple crossovers, chiasmata interference, and because the sum of recombination frequencies between multiple (adjacent) loci can exceed 50%. Chiasmata interference is the phenomena of the suppression of new crossovers in the immediate vicinity of an existing crossover; recombinations do not occur independently of each other (Smith, 1986; Ott, 1991). Complete interference is when multiple crossovers can be excluded between two loci. Mapping functions are designed to transform recombination frequencies into additive map distances (in cM) by incorporating fixed assumptions concerning the number and distibution of multiple crossovers in a defined interval (length); they are tested empirically (Ott, 1991). The mapping function of Haldane assumes that crossovers occur randomly and independently along chromosomes (no interference), contradictory to emperic evidence (Smith, 1986; Haldane, 1916). Kosambi's mapping function, which fits the emperical data fairly well, assumes an intermediate value for the degree of interference (Kosambi, 1944). In general, for small intervals (theta<0.10) the distance (x) between two loci can be assummed to equal the recombination frequency in centimorgans (Morgan's map function): x=theta. By convention, 1cM is considered to represent 1 million bases of contiguous DNA (Donis-Keller *et al.*, 1987). In reality, examination of particular regions or domains of the genome has shown that the actual amount of DNA representing 1cM may vary ten fold in either direction. (Antonarakis *et al.*, 1982; Kazazian *et al.*, 1984).

Linkage Analysis

In theory, any Mendelian trait in humans can be linked to markers and localized. The potential for detecting linkage to a rare dominant disease in a family depends on "the interplay between sample size, degree of polymorphism of marker loci and the recombination distance between the loci in question" (Botstein *et al.*, 1980). Locus heterogeneity, family structure and sizes influence the significance of the results (Lod scores) and the ease at which linkage may be established. Linkage maps determine the expectations of detecting linkage based on the number, spacing and informativeness of the markers used in constructing the map. If one has no prior knowledge of the position of a disease gene, one needs a map which covers the entire human genome with sufficient resolution to detect linkage. In the context of linkage, localizing newly developed markers to a genetic map is analogous to mapping a disease gene. Therefore markers, traits and disease genes are used interchangably in this discussion.

Linkage analysis uses family data to examine the co-inheritance of a disease with polymorphic markers. Under the null hypothesis of independent assortment (no linkage), the theoretical recombination frequency is 50%. This indicates that a disease allele and the marker are on different chromosomes or far enough apart on the same chromosome for them to segregate independently. If a marker is close to a disease gene, they are transmitted together as a block from generation to generation. As the distance between a marker and disease allele decreases, the frequency of recombination between the two also decreases. A recombination fraction significantly less than 50% is evidence for linkage between a marker and the disease. The tightest possible linkage is shown by a recombination fraction equal to zero. As more meiotic events are examined, the significance of the linkage result improves.

Linkage between a disease gene and a marker (or between two markers) is tested using likelihood ratios. The likelihood of a model given a set of observations is defined as the probability that the observations could have occurred given the model. Likelihood ratios are relative values, not absolute ones. They provide a method for comparing models

but do not state that a particular model is correct; only that it is more (or less) favored relative to another model.

The LOD score method of analysis uses affected and unaffected family members to examine the segregation of a disease and markers through consecutive generations. Likelihoods are calculated for linkage with a chosen recombination fraction while specifying the penetrance and mode of inheritance. Penetrance defines the probability of observing a corresponding phenotype for a specified genotype (Ott, 1991). The ratio of this likelihood to that for random assortment (theta=0.5) is the odds for linkage. In linkage analysis each meosis/chromosome is an independent event. Cummulative probabilities are the products of the probabilities for each independent event. By using the log base10 of the likelihood ratios (LOD), the LOD scores from individual pedigrees can be added (assuming no heterogeneity) until significant values are obtained. A cumulative LOD score of >3 is considered significant for demonstrating linkage while a LOD score of <-2 is considered significant for excluding linkage. Once linkage has been detected, the theta value is determined which gives the maximum likelihood for linkage- i.e the highest LOD score.

To be informative in linkage analysis, one parent must be a double heterozygote for the loci being examined. With doubly heterozygous genotypes, the distribution (or phase) of alleles between homologous chromosomes can be in two orientations:

	cis (coupled)		trans (repulsion)	
Chromosome homolog:	1	2	1	2
Locus 1:	A	a	A	a
Locus 2:	B	b	b	B

The phase of a haplotype cannot always be determined. When phase is known, likelihoods are determined directly since recombinants are distinguishable from nonrecombinants. When the phase cannot be determined, likelihoods must be calculated assuming recombinant and nonrecombinant models which are then averaged. This reduces the overall LOD score. The number of families which need to be DNA typed to detect

linkage between markers can only be estimated and is specific for the pedigree structure and number of family members (Botstein *et al.*, 1980). Extended multigeneration families provide the best resource for determining the phase of parental genotypes, and therefore contribute many more informative meioses to a linkage analysis than can be obtained from small nuclear families (White *et al.*, 1985)

An important characteristic of a polymorphism is the degree of allelic variation it displays in the poplulation. Arbitrarily defined, a "polymorphic" locus (as opposed to a "rare" allele) is one in which the most common allele has a population frequency less than 90-95%. Allelic variation must exist between parents at a locus, in order for segregating chromosomes to be distinguished from each other and recombination events detected in their offspring. The likelihood that a mating will be informative is determined by the frequency of heterozygous genotypes in the population based on the number and frequency of alleles at a locus. The chance for detecting linkage between loci increases as the probability of informative matings in the population increases.

One measure of the potential for a marker to be informative is called the Polymorphic Information Content (PIC). PIC is "the probability that a given offspring of a parent carrying a rare allele at a disease locus will allow deduction of the parental genotype at the marker locus" (Botstein *et al.*, 1980). Using this model, the mating frequencies multiplied by the proportion of offspring that will be informative for each mating type are summed determining the PIC.

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

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

The maximum PIC of a 2 allele system is 0.375, while for multiallelic loci the PIC can approach 1.0. As the numbers of alleles increase at a locus, the PIC more closely approximates the % of heterozygous individuals found in the population: Heterozygosity=

$1-\sum p_i^2$. or for a locus of n alleles of equal frequency: H=1-1/n. Thus, the maximum Heterozygosity of a 2 allele system is H=0.5; for 3 alleles H=0.66; for 4 alleles H=0.75 and for 10 alleles H=0.9 In linkage studies a marker with a PIC or Heterozygosity of 0.7 or greater is considered highly informative.

The expected LOD scores vary according to the recombination frequency. Though the full analysis is beyond the scope of this introduction, in a phase unknown situation, a LOD score of 3 can be determined with 300 individuals at theta=0.3 and with as little as 21 individuals from three, 5-sib pedigrees at theta=0 (Botstein *et al.*, 1980). By using multigeneration families this number can be decreased. The maximum contribution to the LOD score in a fully informative situation at theta=0 is 0.3 LOD per individual. A 10 member phase-known sibship would therefore be enough to provide proof of linkage with a significant LOD score of 3.0

The larger the recombination frequency between two loci, the more difficult it is to detect linkage. In small family structures, the expected LOD score falls tenfold as theta increases from 0.1 to 0.3 (Botstein *et al.*, 1980). By placing markers 20 cM apart, a disease gene would lie no more than 10cM from a (informative) marker allowing linkage to be detected. A simple calculation yields: 3300cM/genome ÷ 20cm between markers=165 markers. If the average PIC of markers on the map were 0.5, then 0.5x0.5=.25 would be the probablity that any pair of loci would be informative. If the PIC is increased to 0.7, this probability increases to 0.49 or double the previous value. Greater numbers of high PIC markers will increase the efficiency of linkage maps by increasing the chance that families (matings) will be informative for any pair of markers examined.

We have noted that the ability to detect linkage is based on the structure and number of informative matings in families, as well as the number, informativeness and spacing of markers on the linkage map. The effect of these on the likelihood of detecting linkage for a simple monogenic trait can be calculated relatively straightforwardly. Linkage is much more difficult to detect in complex disorders, where phenotype is the result of locus heterogeneity,

multiple gene interactions, genetic predispositions, incomplete penetrance or environmental factors. Environmental factors such as infection can appear to have a genetic basis in that they cluster and are transmitted through multiple generations. The infectious neurologic disorder Kuru resulting from canibalism among woman and children in New Guinea, is a classic example (Ott, 1991). Some complex disorders confound linkage analysis by introducing uncertainity as to the disease state (genotype) of an individual. The effect of penetrance can be overcome by using only affected individuals in the analysis. To map a disease with a genetic predisposition it is necessary to study kindreds with multiple affected individuals. In these families the disease is most likely to be the result of a single disease gene rather than a chance cluster of sporadic cases with multiple genetic causes. Locus heterogeneity in linkage is usually the result of two types of families being used in the analysis-- those linked to a marker and those unlinked. One approach used in mapping heterogeneous disorders is to find a kindred which is large enough to provide proof of linkage without need for considering other evidence. Another approach is to allow for heterogeneity (admixture) in likelihood models which assume that a certain proportion of families (α) will be linked to a disease at theta (θ) <0.5 and the remainder unlinked at $\theta=0.5$ (Lander & Botstein, 1986). The value of α is not based on external evidence- i.e preclassifying or sorting families into groups on the basis of finer phenotypic classifications (Ott, 1991), rather the analysis calculates the likelihood for linkage and/or homogeneity by varying θ_1 from 0 to 0.5 and the proportion of linked families (α) from 1 to 0 (Lander & Botstein, 1986). $L(\alpha, \theta_1)= \alpha L(\theta_1) + (1-\alpha)L(0.5)$. "Heterogeneity is formulated as H2: $\alpha<1$, $\theta_1<0.5$, and homogeneity is obtained by a single restriction of H2, H1: $\alpha=1$" (Ott, 1991).

Elucidating genetic components of complex disorders is one of the great challanges facing molecular genetics. The increasing informativeness of genetic markers, linkage maps and statistical methods which detect linkage are fundamental to the progress which has been made recently. For example, after 10 years of work, the development of new markers on

chromosome 17 allowed an early onset familial form of breast cancer to be linked to chromosome 17 (Hall *et al.*, 1990) with a LOD score >26 using seven families (MC King, pers. comm).

B: DNA polymorphisms

The human genome displays a great deal of natural variation (polymorphism) between individuals; these have long been recognized by humans as heritable characteristics. They manifest at several levels: in phenotypically obvious characteristics such as eye, hair and skin colour; in molecularly detected protein isoforms and antigens; as heteromorphic regions of chromosomes; and as DNA sequence polymorphisms. It is thought that approximately 1 nucleotide in 200 may vary between people as demonstrated by DNA sequencing (Antonorakis *et al.*, 1982; Jeffreys *et al.*, 1988). This tremendous source of polymorphism can be exploited by the ability to characterize DNA molecularly, and has resulted in new classes of genetic markers. These usually detect DNA polymorphisms (insertions, deletions or single base changes) which alter the electrophoretic mobility of DNA; they vary considerably in informativeness and ease of use.

Restriction fragment length polymorphisms (RFLP) are of historic importance in linkage map development since they were the first type of polymorphism demonstrated to be abundant and ubiquitously distributed throughout an organism's genome (Botstein *et al.*, 1980; White *et al.*, 1985). Over 3000 thousand have been described and used in genetic linkage studies (Litt, 1991). Most RFLPs are the result of single nucleotide changes in restriction enzyme recognition sequences (site polymorphisms). These alter the length of restricted DNA fragments due to the presence or absence of an enzyme cut site. Alternatively, RFLP's can result from variable numbers of tandem repeats (VNTRs) being contained within a restriction fragment.

An individual's genomic DNA is digested with a restriction enzyme, fractionated by electrophoresis and transfered to membranes using the technique of Southern. Southern

blots are probed with single copy DNA sequences which hybridize to complementary genomic sequences present on the membrane and reveal any alterations of restriction fragments due to an RFLP. Generally, RFLPs are 2 allele systems, of limited informativeness, which gain power upon haplotyping.

Loci consisting of repetive elements have proved to be valuable genetic markers. There are several loci known where the number of repeats within a block of tandemly repeated DNA varies among individuals. These include alpha satellite DNA (Warburton *et al.*, 1991), ALU associated polyA sequences (Economou *et al.* 1990), minisatellites/VNTRs (Jeffreys *et al.*, 1985a; Nakamura *et al.*, 1987) and short tandem repeats (STR) (Weber & May, 1989; Litt & Luty, 1989). Of these, VNTRs and STRs are most widely used as genetic markers.

Variable number of tandem repeat (VNTR) polymorphisms consist of sequences 2-60 nucleotides in length repeated 10's-1000's of times. Traditionally they are revealed by probing Southern blots of restricted genomic DNA with single copy probes near the VNTRs. Nakamura showed that they are multiallelic and often very informative in genetic linkage studies (Nakamura *et al.*, 1987). Minisatellites are multilocus systems because the core repeat sequences are present at different locations in the genome (Jeffreys *et al.*, 1985b,1988). Minisatellites, visualized with core sequence probes on Southern blots, have been popularized as DNA fingerprints and are used in forensic and genetic identity studies (Hagelberg *et all.*, 1991). Though much more informative than RFLP site polymorphisms, they are found predominately in telomeric regions of chromosomes which limits their usefulness in mapping studies (Litt *et al.*, 1990.- see Appendix I.a).

In addition to RFLPs and VNTRs, new approaches have been developed to detect single nucleotide polymorphisms, though these have been much less important for linkage map constructions than RFLP's and VNTRs. Partial mismatching between complementary DNAs allows the mismatched nucleotide(s) to be chemically or enzymatically modified, cleaved and detected as charcteristic fragments on polyacrylamide gel electrophoresis

(PAGE) (Cotton *et al.*, 1988; Meyers *et al.*, 1985). DNA melting temperature (Tm) differences resulting from mispairing can be distinguished as heteroduplexes on denaturing gradient gel electrophoresis (DGGE) (Meyers & Maniatis, 1986). A nucleotide substitution can alter the secondary conformation of single stranded DNA. Single strand conformational polymorphisms (SSCP) are detected as altered mobilities of DNA strands in nondenaturing PAGE (Orita *et al.*, 1989). Ideally, each method is specific, characteristic of the mutation and experimentally reproducible. In reality, these techniques have found limited use in genetic linkage studies because they are technologically demanding and not particularily efficient or informative. They are mostly used to scan for polymorphisms or mutations in a limited number of sequences, such as candidate disease genes or coding sequences.

Microsatellites

Microsatellites, variously called short tandem repeats (STR) and simple sequence repeats polymorphisms (SSRP) are highly repetitive interspersed DNA elements found in all eukaryote genomes examined to date (Tautz, 1989; Stallings *et al.*, 1991). They consist of 2-8 deoxyribonucleotide sequences tandemly repeated 5-60 times. The number of repeat units at a locus can vary between individuals, resulting in DNA length polymorphisms of the VNTR class. Microsatellites typically display many alleles (1->15) and allow individuals to be efficiently DNA typed by PCR. Depending on the organism and the type of repeat, there are hundreds to tens of thousands of microsatellite loci per genome (Levinson & Gutman, 1987; Tautz, 1989; Stallings *et al.*, 1991). (CA)n dinucleotide repeats are the most abundant STR; they have revolutionized genetic map construction by providing a class of polymorphism which is abundant, ubiquitously distributed throughout the genome, highly informative and easy to clone and use. Since CA microsatellites were first found to be polymorphic and inherited in a Mendelian fashion by Weber & May (1989) and Litt & Luty (1989), thousands have been characterized for genetic map

constructions of many different genomes, including human, mouse, dog and pig (Hazan *et al..*, 1992; Kwiatkowski *et al.*, 1992; Ostrander *et al.*, 1992; Wintero *et al.*, 1992). Since the discovery of CA microsatellites, other short repeat motifs of 3-6 nucleotides have been examined and found to be very polymorphic (Edwards *et al.*, 1991a). However, the frequency of high PIC markers per non CA STR motif or class is not known, and these repeats are less abundant than CA microsatellites in the human genome. (See Discussion).

The abundance, even distribution and often high PIC of CA repeats make them the most useful polymorphism for genetic mapping. CA repeats were demonstrated to be randomly distributed over the entire genome by the high frequency of repeat containing clones in well characterized total human genomic libraries. Based on the fraction of CA microsatellite containing clones in the libraries, it has been estimated that there are upwards of 50,000 CA repeats in the human genome (Hamada & Kakunaga, 1982; Tautz, 1989; Weber & May, 1990). Microsatellite sequences do not appear to be underepresented in cloned libraries, unlike large VNTR repeats which may be inherently less amenable to cloning.

In 1988, Litt and Luty amplified by PCR a microsatellite contained in the 4th intron of the cardiac actin gene (Litt & Luty, 1989). Independently, Weber observed that the number of CA repeats varied in GeneBank data enteries for identical sequences from different laboratories (Weber & May, 1989). Both of these studies showed that CA microsatellites are highly polymorphic and inherited in a Mendelian fashion.

Figure 1 shows a DNA sequencing autoradiograph of 4 typical STRs from different loci which are polymorphic because of variation in the length of the repeat. Other alleles at those loci have different numbers of repeats surrounded by identical flanking sequences.

Figure 2 shows how alleles are visualized by the polymerase chain reaction. The arrows represent PCR primer sequences that are used to amplify genomic DNA from individuals. Though there are thousands of repeats in the human genome, the specificity of

Figure 1: DNA sequencing autoradiograph of 4 microsatellites. The DNA sequences for D14S43 and D9S55 transcribed from the autoradiograph are shown in figures 12 & 17A.

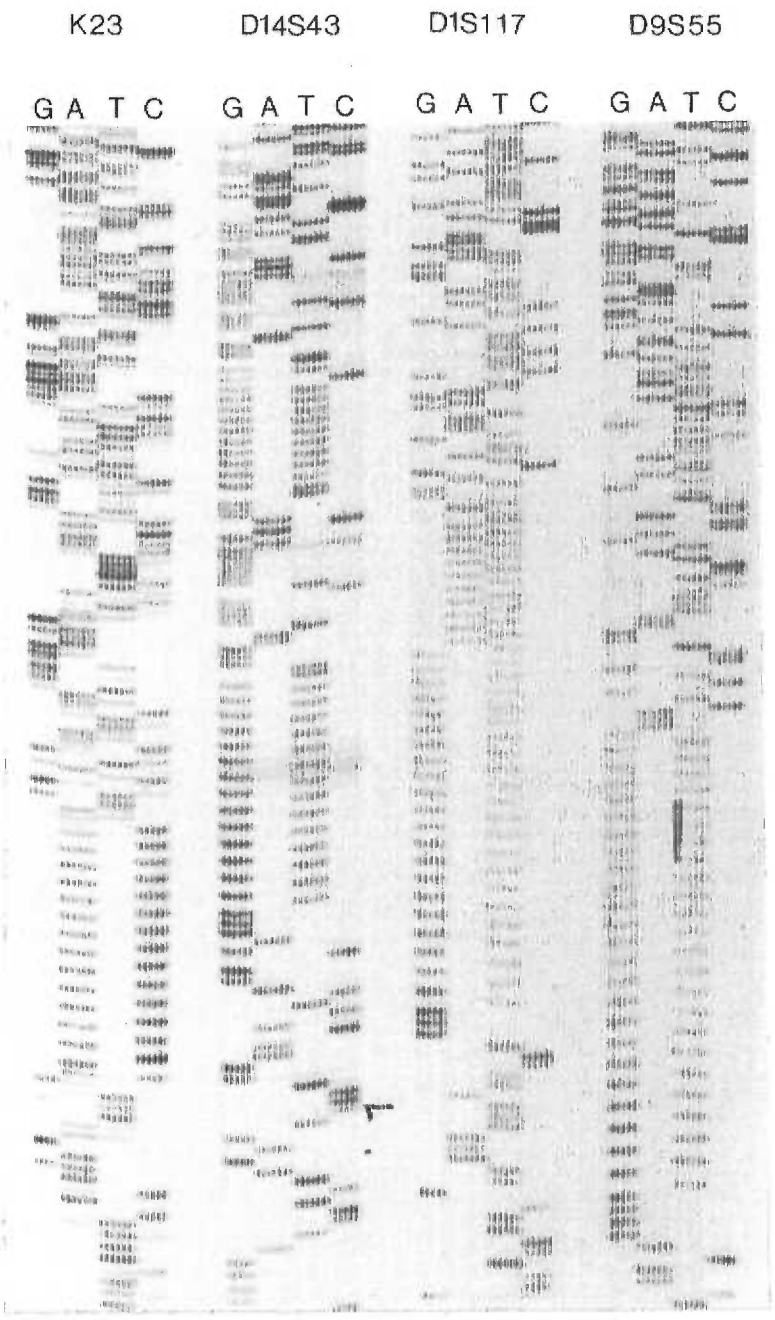

K23
D14S43
D1S117
D9S55
G A T C
G A T C
G A T C
G A T C

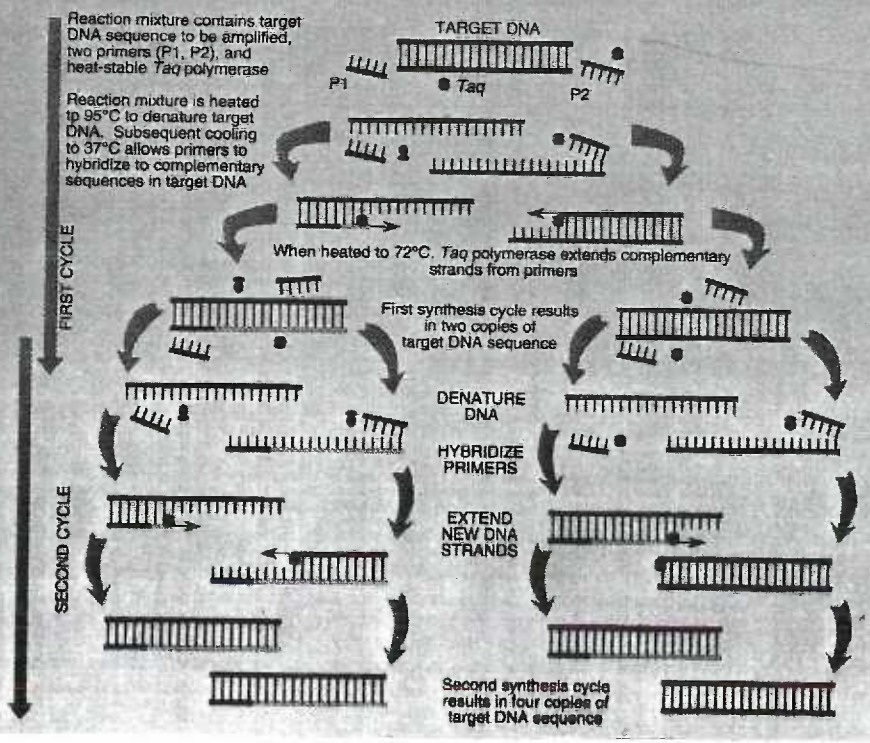
Figure 2: Cartoon representation of the polymerase chain reaction (PCR) (from Micklos & Freyer, 1990).

PCR (in vitro DNA amplification)

Described as being to genes what Gutenberg's printing press was to the written word, PCR can amplify a desired DNA sequence of any origin (virus, bacteria, plant, or human) hundreds of millions of times in a matter of hours, a task that would have required several days with recombinant technology. PCR is especially valuable because the reaction is highly specific, easily automated, and capable of amplifying minute amounts of sample. For these reasons, PCR has also had a major impact on clinical medicine, genetic disease diagnostics, forensic science, and evolutionary biology.

PCR is a process based on a specialized polymerase enzyme, which can synthesize a complementary strand to a given DNA strand in a mixture containing the 4 DNA bases and 2 DNA fragments (primers, each about 20 bases long) flanking the target sequence. The mixture is heated to separate the strands of double-stranded DNA containing the target sequence and then cooled to allow (1) the primers to find and bind to their complementary sequences on the separated strands and (2) the polymerase to extend the primers into new complementary strands. Repeated heating and cooling cycles multiply the target DNA exponentially, since each new double strand separates to become two templates for further synthesis. In about 1 hour, 20 PCR cycles can amplify the target by a millionfold.

DNA Amplification Using PCR



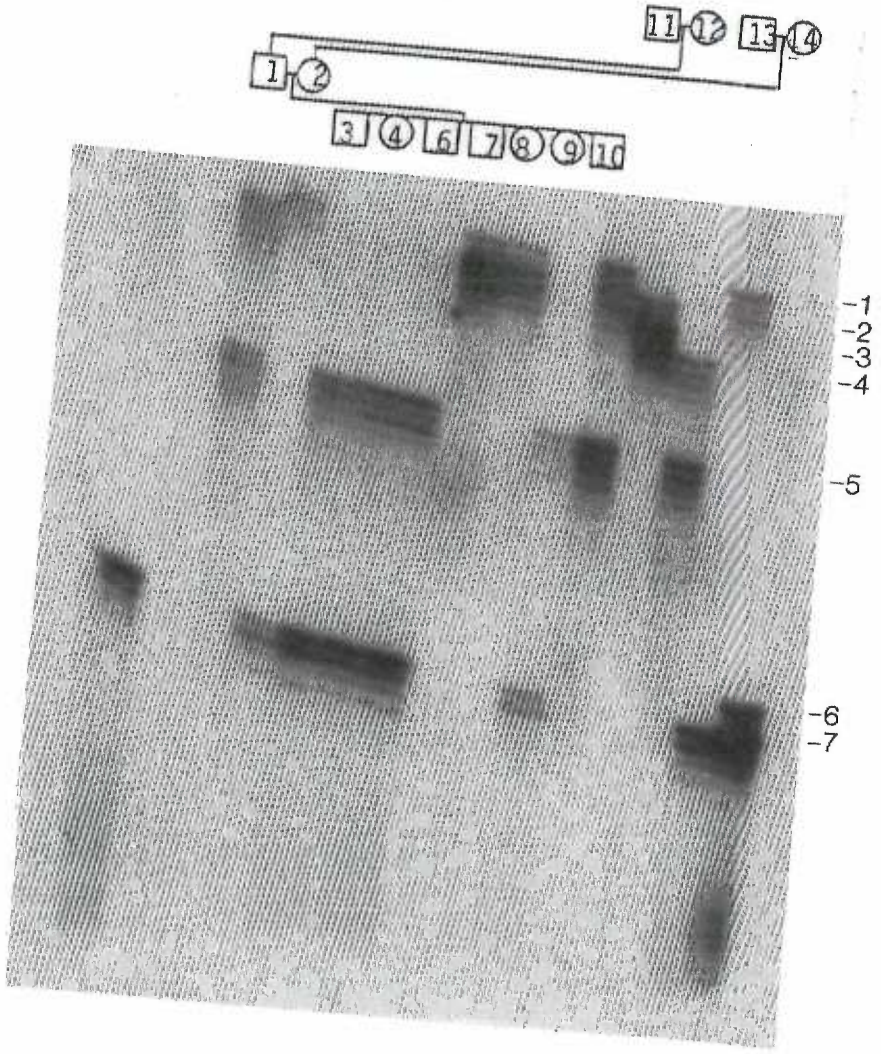
the PCR is based on the unique single copy sequences which flank repeats. Utilizing thermal stable DNA polymerases which withstand the high temperatures required for disassociating DNA, alternate cycles of primer annealing, primer extension and DNA denaturing allows DNA sequences between primers to be exponentially amplified. The product of the PCR is elecrophoresed on denaturing polyacrylamide gels; alleles which vary in length by as little as a single nucleotide can be resolved and distinguished from each other.

A major advantage of microsatellites is that many of them have high PICs. Weber's original report examined 10 repeats; 5 were sequences in GenBank and 5 were from phage isolated from a human chromosme19 genomic library (Weber & May, 1989). All ten were polymorphic with alleles varying by increments of two nucleotides. The number of alleles and PIC ranged from 5-11 and 0.31-0.79 respectively. A recent summary of 51 repeats developed in our laboratory from 1989-1992, showed that 23 (45%) have PICs>.70 and 37 (72%) have PICs>0.5 . The median PIC for all 51 repeats, including nine monomorphic loci, is 0.65 (Litt, pers. comm.).

The high degree of informativeness of CA microsatellites is demonstrated in Figure 3. A three generation pedigree shows 7 alleles segregating within it. The parents are fully informative and the genotypes of all the children are easily determined directly from autoradiographs. For comparison, when allele frequencies are equal, the maximum heterozygosity for a two allele system is 50% which corresponds to a PIC 0f 0.37. Since allele frequencies are usually not equal, most RFLPs have PIC's less than 0.3. To increase the informativeness of RFLP markers, haplotypes can be constructed between closely linked markers. However, to benefit linkage analysis the phase of the haplotype on a chromosme must be determined. Often for multiply heterozygous individuals phase cannot be determined without knowing the parental genotypes. This is a serious drawback when studying small families or late onset disorders where parents may be deceased and

Figure 3: The genotypes of CEPH family 1423 at the D14S43 locus. There are seven alleles in this family. Father is heterozygous for alleles 2 & 5 and mother is heterozygous for alleles 1 & 7; alleles 3, 4, & 6 are present only in the grandparents. In individuals 7, 8, 11, 14, alleles which differ by a single repeat unit (GT) are clearly distinguishable and scorable on the basis of relative band intensities.

D14S43: CEPH K1423



unavailable for DNA typing. The high PIC of microsatellites are particularly valuable for proving or disproving linkage by the affected sib pair (ASP) method, which does not require participation of other family members. Sib pair analysis establishes linkage based on the frequency of shared alleles of high PIC loci (or haplotypes of compound loci) between affected sibs, compared to random segregation expectations (Goldin & Gershon, 1988; Payami 1984).

There are many speculations about the origin of microsatellites, their possible function(s), and the basis for allelic diversity. The position of microsatellites have been found to be highly conserved in closely related species (Moore *et al.*, 1991). The presence of a number

of conserved microsatellites between highly divergent species such as rat and man suggests that some loci have ancient origins (Stallings *et al.*, 1991). *In vitro*, alternating pyrimidine/purine nucleotide stretches are capable of assuming left handed Z DNA conformations instead of the more usual B form (Holliday, 1989). Z DNA conformations have relatively low melting temperatures (Tm), relieve negative supercoiling stress in plasmids, and promote the formation of triple helical domains (Baran *et al.*, 1991). These are properties of DNA which may mediate gene expression or chromatin structure in unknown ways. For exmple (CA)n sequences have been shown to enhance the transcriptional activity of genes in plasmid constructs, and form sequences which are spontaneouis deletion hot spots (Freund *et al.*, 1989). Mini- and microsatellites have been hypothesized to be hot spots for recombination between homologous chromosomes. Unequal crossing over during meiosis was proposed to account for the creation of new alleles in minisatellites (Jeffreys *et al.*, 1988), but subsequent studies of new mini and microsatellite alleles showed that flanking genetic markers were not exchanged, as would be expected in crossing over (Wolff *et al.*, 1989; Morral *et al.*, 1991). Also, new minisatellite mutations have been observed in somatic cells independent of meiosis (Armour *et al.* 1989). Still, microsatellites may facilitate gene conversion, unequal sister chromatid

exchange or other mechanisms which insert or delete repeat units. One study has documented gene conversion events in sequences flanking STRs (Sleightom *et al.*, 1980).

In order to explain how new microsatellite loci and alleles arise, Levinson and Gutman have proposed a model based on the slipped strand mispairing of STRs during replication and repair of DNA (Levinson & Gutman, 1987). The central tenet is that new loci and allelic variants arise out of core repeat sequences which reach "threshold" sizes by chance, after which they become self-expanding (polymorphic) due to unfaithful replication and/or repair of repeat DNA. This suggests that the abundance of each type of repeat is related to the number of potential (expandable) core sequences present in the genome and is not dependent on transposition or DNA exchange/crossover events.

C: Chromosome 21

Chromosome 21, because of important genetic diseases and it's small size, is serving as a prototype for human genome studies by providing fundamental information regarding chromosome organization at all levels of resolution. Since it is the smallest human autosome, it should be the easiest to map completely genetically and physically. Physical characterizations include the construction of detailed physical maps based on pulse field gel electrophoresis of restriction fragments (Gardiner *et al.*, 1990ab; Owen *et al.*, 1990; Ichikawa *et al.*, 1992), YAC and cosmid contig assembly (McKormick *et al.*, 1989), microdissection (Yu *et al.*, 1992), cloning single copy sequences and genes (Neve *et al.*, 1986; Stewert & Kurnit, 1990), integrated physical, transcriptional and linkage mapping (Nizetic *et al.*, 1991), radiation hybrid mapping (Cox *et al.*, 1990), identifying CpG clusters, recombination hot spots, gene distributions, and defining breakpoints of frequent chromosome 21 translocations and rearrangements (Patterson *et al.*, 1990; Gardiner *et al.*, 1990ab). As genetic maps become more complete, much will be learned concerning the large scale sequence organization of chromosomes, "gene environments" and human disease. Questions which can be addressed include: "What relationship, if any, does linear

order have with genetic regulation? with gene interactions ? with position effect? How much redundancy is there in genetic information? What effect do heterochromatic regions have on the genetic material of euchromatic segments? What is the significance of finding sex differences in recombination in specific chromosomal regions? What is the relationship between physical and genetic maps?" (Conneally & Rivas, 1980). Comparative mapping studies examining the evolution, origin of chromosomes (Threadgill *et al.*, 1991) and the nature of conserved syntenies across multiple species may be useful for positioning human disease genes.

There are a number of important genetic conditions which map to chromosome 21. Down's syndome (DS) is the most common form of mental retardation and the result of trisomy 21. Cytogenetic and molecular studies have examined the parental origin of the extra chromosome in order to understand the basis for nondisjunction in trisomy 21 (Antonorakis *et al.*, 1991). Recent studies of individuals with partial trisomy 21 have identified minimal regions of chromosome 21 which are involved in the pathogenesis of various features of DS (Korenberg *et al.*, 1990). Efforts are underway to clone genes responsible for familial amyotrophic lateral sclerosis (Siddique *et al.*, 1991) and progressive myoclonus epilepsy which have been linked to chromosome 21. These studies, and others which seek to prove (or disprove) linkage of a subset of familial Alzheimer's disease (FAD) to chromosome 21 will benefit by the development of polymorphic loci which improve the genetic linkage map of this chromosome.

Chromosome 21 contains approximately 1.7% of the human genome or 60 megabases (Mb) of DNA. Of this, approximately 43 Mb are in the long arm. The most comprehensive physical mapping study of chromosome 21 is by Gardiner *et al.*, 1990ab. Using pulsed-field gel electrophoresis, Gardiner has ordered a collection of 54 single copy sequences on 33 Not1 restriction fragments, accounting for 43 Mb of DNA or virtually all of 21q. Gardiner reveals that 60% of the unique sequence probes are in 11 physical linkage groups, which are contained on only 20% of the long arm. There are twice as many

unique sequences near the distal end of chromosome 21 than on the proximal part; 85% of transcribed sequences and genes are within q22; 62% (13/21) are in q22.3 which contains only 15% of chromosome 21's DNA.

In 1989 when this study began, the proceedings of the 10th Human Genome Mapping Workshop (HGM 10) listed 34 expressed sequences and 123 anonymous D segments, which between them accounted for 54 polymorphic loci (Kidd *et al.*, 1989; Carritt & Litt, 1989). Only 2 of the polymorphisms had PICs greater than 0.50. Compiled in 1991, HGM 11 lists 201 D segments, 176 of which are regionally defined; and 38 expressed sequence, 18 of which represent cloned genes (Cox & Shimizu, 1991). There are 21 polymorphic markers which have PIC's greater than 0.5. Only ten have been mapped to specific regions of chromosome 21: 5 of which were developed in this dissertation.

A number of linkage maps have been constructed of chromosome 21 (Tanzi *et al.*, 1990; Warren *et al.*, 1989; Donnis-Keller *et al.*, 1987). Warren *et al.l* ordered 17 markers on chromosome 21 which covered 136 cM when averaged between the sexes. An updated version of this map (Petersen *et al.*, 1991) is the most comprehensive linkage map of chromosme 21 (Figure 4). Forty CEPH families were typed with 27 markers yielding a sex-averaged map length of 158 cM, with an average spacing of 7 cM between markers. Twenty five of 27 markers used by Petersen were RFLPs; 2 were CA microsatellite polymorphisms. The observed heterozygosity of the loci used in the map averaged 0.45 and ranged from 0.11 to 0.76. This map covers most of the long arm of chromosome 21; 50% of the genetic distance is in q22.3 which is less than 15% of the physical length of chromosome 21. The most proximal marker on the map is D21S13/16; only 3 loci (D21S13/16, D21S110 & D21S1/11) were available to define the proximal half of 21q (21cen-q21.2). The most distal genetic marker, COL6A1, is estimated to be 600 Kb from the telomere (Burmeister *et al.*, 1991). There is a nine fold difference in the frequency of recombination between males and females in the region of 21q21 (Cox & Shimizu, 1991).

Figure 4: A linkage map of Chromosome 21 showing the genetic distances between commonly used polymorphic markers.(from Petersen *et al.*)

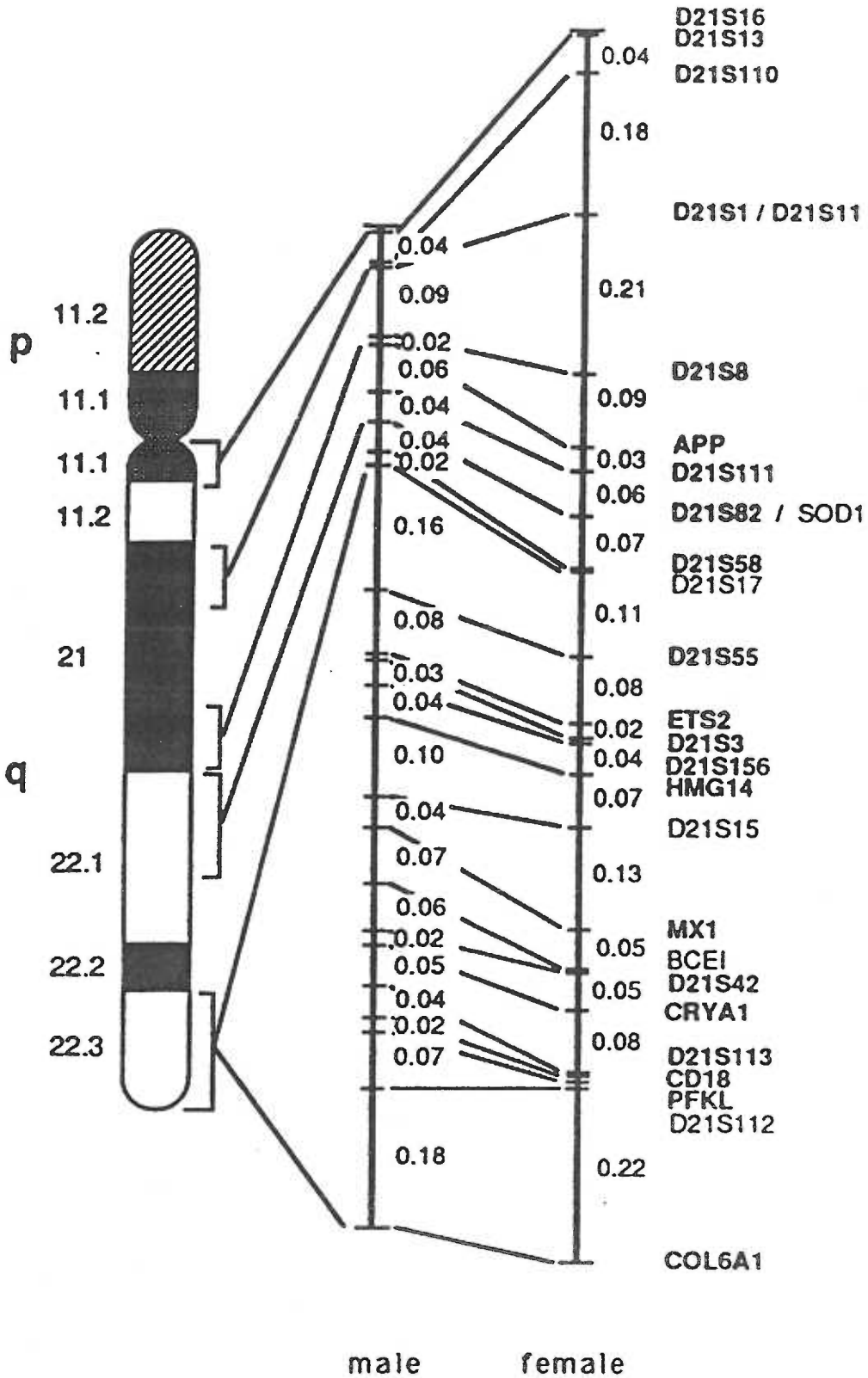


FIG. 4. Genetic linkage map of human chromosome 21 in males and females (Haldane map function, variable ratio in recombination between males and females). Recombination values between loci are indicated. Physical locations are shown for selected markers.

The interval 21q11.2-cen remains of interest for physical mapping studies and genetic marker development. D21S13/S120 represents the proximal end of the genetic and physical map (Butler *et al.*, 1992; Gardiner *et al.*, 1990ab). The interval between these markers and the centromere is one of the few remaining gaps (2 Mb) on the physical map of chromosome 21 (Cox & Shimizu, 1991). The genetic size of this region is unclear because of the scarcity of markers in the region. This area has fewer single copy sequences than other regions of chromosome 21; band 21q11.2 is Geimsa negative and thus more likely to contain expressed genes than adjacent bands (Butler *et al.*, 1992). The D21S13 region remains implicated in the etiology of one form of familial Alzheimer's disease (FAD). We have made an effort to develop markers from the proximal half of the long arm of chromosome 21, an interval defined cytogenetically by 21q21-cen and genetically by the markers D21S13/S120 and D21S1/S11.

Familial Alzheimer's disease (FAD)

Alzheimer's disease (AD) is a degenerative disorder of the central nervous system and the leading cause of severe and progressive dementia in the western world. It is estimated that 10 % of the U.S. population over 65 years and 47 % over 85 years have Alzheimer's disease (Evans *et al.*, 1989). The etiology of AD is unknown but environmental and genetic factors are suspected of being contributory to the disease (St. George- Hyslop *et al.*, 1987a). About 5-10% of Alzheimer's disease appears to segregate in families as an autosomal dominant disorder (denoted FAD) and has been the object of many genetic linkage studies. The possibility of chromosome 21 containing a gene locus involved in the etiology of AD arose from observations that people with Down's syndrome (trisomy 21) have some of the same phenotypic and pathophysiologic traits as people affected with AD. Both conditions involve progressive dementia which correlate with neurofibrillary tangles and senile plaques (Tanzi *et al.*, 1991; Bird *et al.*, 1989). The substantial deposition of beta 4 amyloid in the brains of people with DS and AD suggests

that it is important in the pathogenesis of AD. The localization to chromosome 21 of the amyloid precursor protein (APP)- from which beta 4 amyloid is processed, suggested that APP may underlie the genetics of AD. The first linkage studies of FAD were done using polymorphisms in the obligate Down's syndrome region (21q22) and the cloned APP gene. The initial excitement about the genetic involvement of APP subsided when recombination was detected between the gene and FAD in these families suggesting that it was not at the site of the inherited defect underlying this disorder (Tanzi *et al.*, 1987).

The first study to show linkage between chromosome 21 and FAD implicated a region near two loci- D21S1/S11 & D21S16 (St. George Hyslop *et al.*, 1987). This interval spans the proximal half of band 21q21 through 21q11.2. A study by Goate *et al* also suggested that D21S16 and D21S13 were linked to FAD in 6 families (Goate *et al.*, 1989). However, in a study using a group of families thought to display a monogenic variant of FAD, Schellenberg failed to find linkage of FAD to some of the same markers (Schellenberg *et al.*, 1988). Pericak-Vance *et al* (1990), used the affected pedigree member method to study predominately late onset FAD individuals and found no linkage of FAD to chromosome 21. A comphrensive re-examination of the linkage data available in the literature by St. George-Hyslop *et al* ., appears to confirm linkage of early onset FAD to markers D21S13/S16 and D21S1/S11. Pairwise analysis gave a LOD score of 3.6 at theta=0.17 with D21S13/S16; and LOD= 4.5 at theta= 0.17, with D21S1/S11. This study revealed that only a subset of families contributed to the positive LOD score. Multipoint linkage aimed at localizing the defect relative to the two linked loci resulted in peak LOD score>5.0 for two positions of chromosome 21: one proximal to D21S13/16 and the other distal to D21S1/11 (St. George-Hyslop *et al.*, 1990).

Recently a specific mutation in the APP gene has been found which appears to segregate in affected individuals in two families (Goate *et al.*, 1991). This mutation is a nucleotide substitution which changes a valine to isoleucine at amino acid position 717 in exon 17 and has subsequently been described in additional families (Fidani *et al.*, 1992).

While mutations in APP may cause AD in families, they are exceedingly rare and unlikely to be a major cause of FAD (Kamino *et al.*, 1992; Tanzi *et al.*, 1992).

The genetics of AD remains unclear. Uncertain diagnosis, genetic heterogeneity and late age of onset are confounding issues in AD linkage studies. Only 5-10% of Alzheimer's disease appears to segregate in families, making their identification and collection difficult. The late age of onset of the disease makes gathering data difficult, since parents and grandparents of affected people are usually deceased, and younger family members cannot be diagnosed. A diagnosis of AD does not necessarily insure the elimination of confounding effects due to genetic heterogeneity of potential subtypes of AD. In linkage analysis, inclusion of families which are non allelic for the disorder, can "cancel out" a positive contribution to the LOD score by a family which is linked. In an effort to improve the chances for detecting linkage between FAD and genetic markers, criteria have been sought which attempt to distinguish between subtypes of FAD. These include the age of onset of AD, sex, epidemiological, clinical and pathological variables. Because of the difficulty of diagnosis of AD, most current linkage studies seek individuals and families which have a positive family history for the disease in multiple generations, contain multiple affected sibs and an autopsy confirmed diagnosis of AD in a deceased family member. In summary, evidence for genetic heterogeneity in AD includes the demonstration of a mutation in the APP gene which may be causative of FAD (Goate *et al.*, 1991), linkage of early onset FAD to chromosome 21 markers in the most proximal region of 21q (St. George-Hyslop *et al.*,1990; Goate *et al.*, 1989), the lack of linkage to chromosome 21 in other FAD families (Schellenberg *et al.*, 1988,1990) and the tentative linking of late onset FAD to chromosome 19 (Pericek-Vance *et al.*, 1990).

D: Proposal: Specific aims and significance.

Most current genetic markers on linkage maps suffer from three limitations: 1) Mediocre informativeness as measured by heterozygosity or PIC; 2) Non random

distribution of markers resulting in clusters or gaps on maps; and 3) The use of obsolete technologies for genotype determination. We think that microsatellites/(CA)n dinucleotide repeats are the best type of marker for constructing genetic maps because of their abundance, informativeness and ease of use. By cloning and mapping by linkage analysis a set of highly informative microsatellites derived from chromosome 21, a genetic linkage map will be constructed. The markers and mapping data will contribute to the US Human Genome Initiative Index Map and the Center for the Study of Human Polymorphisms (CEPH) consortium map currently under construction and will facilitate the cloning of disease genes from physically and genetically defined sub-regions of chromosome 21, once linkage has been firmly established. In addition, microsatellites will be developed from the proximal half of 21q, defined by the cytogenetic interval 21cen-q21, and by the genetic markers D21S13/S16 and D21S1/S11. Gardiner's study , the compilitations of HGM 10 &11 and the limited number of probes used in constructing published linkage maps are evidence that significant portions of 21q remain poorly mapped. HGM11 only lists 3 markers from this region with PICs>.0.5. Published genetic maps have not been extended proximally from D21S13 where a 2 Mb gap remains between the centromere and D21S13/S120. This region is implicated in the etiology of one form of familial Alzheimer's disease. "Continued linkage analysis of FAD pedigrees with highly informative polymorphisms (e.g. simple sequence repeats) from chromosome 21 loci should resolve the issue of whether a familial Alzheimer's disease gene defect distinct from APP is located on chromosome 21" (Tanzi *et al.* 1991). In particular, microsatellites from landmark loci D21S1/S11 and D21S13/S16 (which span a putative FAD locus) will be developed. In addition to improving the PIC of these loci for FAD linkage studies, developing microsatellite polymorphisms will allow individuals and kindreds to be efficiently genotyped by PCR (see Discussion /Kamino *et al.*, 1992). Finally, in the course of this study we expect to develop technology improvements for cloning and characterizing microsatellite repeats.

II: MATERIAL & METHODS

A: Flow chart for characterizing CA repeats

Chromosome specific cosmids
Screen with (CA)n probe

CA positive cosmids
Southern blots >>>>>> Size and number of repeat containing inserts
Subcloning

CA positive subclones
DNA sequencing >>>>>> PCR primers

PCR/Locus Characterization
CEPH parents >>>>>> # of alleles/% Heterozygosity

Cell hybrid panel >>>>>> chromosome of origin

CEPH families >>>>>> Mendelian Inheritance
>>>>>> 2 point linkage analysis >>>> rough localization
>>>>>> multipoint linkage analysis >>>> fine mapping

B: Materials

Reagents: Restriction enzymes, T4 polynucleotide kinase, T4 DNA ligase, DNA polymerase I, Klenow fragment, alkaline phosphatases were obtained from the following sources: Bethesda Research Laboratories, New England Biolabs, United States Biochemical Corp. and Boehringer Mannheim. DNA polymerases used in PCR (AmpliTaq & Replinase) were from Perkin Elmer/Cetus and Du Pont respectively. DNA sequencing kits (Sequenase V2.0) were from US Biochemical Corp. DNA preparation and purification columns were from Qiagen Corp. [Gamma^{32}P]-ATP and [alpha^{32}P]-dCTP were from Amersham and New England Nuclear. The plasmid cloning vector, pTZ18U was

from US Biochemical Corp. CA/GT copolymer was from Pharmacia. Nylon and nitrocellulose membranes used in hybridization protocols were from Amershan (Hybond N+) and Schleicher and Schuell (Nytran and nitrocelluose). Oligoprimers used in PCR were from Research Genetics (Huntsville, AL) or synthesized in our laboratory on an Applied Biosystems PCRmate™.

Probes: Single copy probes for D21S8 (pPW245d), D21S1 (pPW228c) and D21S11 (pPW236b) were provided by Dr. Paul Watkins (Watkins); D21S16 (NH2) by Dr. G. Stewart and D21S13 (G21RK) by the American Type Tissue Collection. Clones provided as bacterial stabs or lyophilized phage were grown, DNA prepared and appropriate inserts isolated for use as probes.

Libraries: Cosmid clones were from two different libraries constructed from flow sorted human chromosome 21 DNA. A nylon membrane (22cm x 22cm) consisting of DNA from 9216 arrayed cosmid clones from Library# ICRFc102 was provided to us by H. Lehrach (Nizetic *et al.*, 1991). Additionally, we received two 96 well titer plates with frozen stocks of random clones from this library. Library# LL21NC01 was constructed by P. de Jong at the Lawrence Livermore National Laboratory (de Jong *et al.*, 1989). Seventeen cosmids mapped by *in situ* hybridization to human chromosome 21 and 384 unmapped cosmid clones were provided to us as stabs and frozen stocks.

Reference families: Human genomic DNA from 40 two- or three-generation families were provided by the Center for the Study of Human Polymorphism (CEPH). These consist of 560 individuals from 27 Utah Morman, 10 French, 2 Venezuelan and I Amish families (Dausset *et al.*, 1990).

C: Screening & Cloning Repeats

Identification of repeat bearing clones: DNA preparation, subcloning, labeling reactions, single copy probings and Southern blot analysis followed the protocols outlined in Maniatis, Fritsch and Sambrook (Maniatis *et al.*, 1982). Cosmid clones in 96 well titer plates were replica plated onto nitrocellulose filters and grown overnight at 37°C on LB agar/kanamycin (30ug/ml) plates. Grown bacterial colonies from subclonings were lifted from LB/ampicillin plates (100ug/ml) onto nitrocellulose. Colonies were denatured, neutralized and fixed on to membranes by alkaline treatment and baking for 1.5 hrs at 80C. CA repeat containing clones were visualized by probing colony lifts, Southern blots and the 96x96 array with CA/GT copolymer or $(CA)_{15}$ oligomer. CA/GT copolymer and single copy inserts were labeled with [alpha ^{32}P]- dCTP by nick-translation or random oligomer priming of DNA with Klenow fragment of DNA polymerase 1. Oligonucleotides and the $(CA)_{15}$ oligomer were 5' end labeled with [gamma ^{32}P]-ATP using T4 polynucleotide kinase. The radioactivity incorporated in labeling reactions was determined by trichloroacetic acid precipitations and scintilation counting. Hybridization conditions were as follows:

Copolymer probe: Filters containing colony or plaque lifts were screened with 10^6 counts/ml nick-translated poly (dT-dG).poly(dA-dC) probe. Hybridizations were performed overnight at 65°C in 0.5 M sodium phosphate pH 7.0, 7% (w/v) SDS, 1% (w/v) BSA (Church *et al.*, 1988, 1991). Filters were washed at a maximum stringency of 0.1X SSC/0.1% SDS at 55°C and autoradiographed.

$(CA)_{15}$ and other oligomer probes: Filters were hybridized 1-2 hrs. at 60°C with 1-2 pmol/ml of end labeled oligomer in 5XSSPE/0.1% SDS, washed twice in 6XSSC/0.1% SDS at 60C for 15-30 minutes and autoradiographed.

Oligonucleotide pools: Filters were prehybridized in 3 M tetramethyl ammonium chrloride, 50 mM sodium phosphate pH 6.8, 1 mM EDTA, 0.5% SDS, 5X Denhardt's solution, 100 µg sheared salmon sperm DNA for 2 hours at 65°C. Filters were hybridized

overnight in the same solution plus 2 pmol/ml of end-labeled oligonucleotides at 60°C. Filters were washed once in 6XSSC/0.1% SDS for 10 minutes at room temperature and then twice for 20 minutes in TMACl wash solution at 60°C and autoradiographed (Wood *et al.*, 1986).

Subcloning and DNA sequencing of repeats for design of PCR primers: DNA from 3 ml overnight cultures of cosmids containing a CA repeat was digested with the restriction enzymes, Sau3A, AluI, HaeIII and AluI/HaeIII and Southern blotted from 1.5% agarose gels to nitrocellulose or nylon membranes. Blots were hybridized to nick translated copolymer or end labeled $(CA)_{15}$ oligo, in order to display the size of repeat bearing restriction fragments. Digests displaying repeats on fragments of 50-550 bp were subcloned into the plasmid cloning vector, pTZ18u. Ligations were with 50 ng of dephosporylated vector, 1-5 ng of restriction digested DNA, 0.5 units of T4 DNA ligase in 10 ul rxns incubated overnight at 15C. One quarter to one third of the ligation mixture was used to transform 100 ul of DH5alpha frozen competent cells per Hannahan . Colonies were lifted and rescreened with probes to identify subclones with repeats.

Plasmid DNA for DNA sequencing was prepared from 3 ml overnight bacterial cultures using Qiagen 20 columns. Double stranded templates were sequenced by the dideoxy chain termination method using Sequenase™ and reagents provided in kit form by US Biochemical Corp.

D: PCR

Primer Design: Sequences for oligonucleotides 17-25 bases in length were choosen from repeat flanking sequences which are predicted to give a PCR product of 90-180 bp. Primer sequences were chosen so as to not complement each other and to have predicted melting temperatures (Tm) that were within 2°C of each other. Predicted annealing/melting temperatures for primers were calculated using the formula: 4°C(G + C nts.)+ 2°C(A +

T nts.)= Tm . Some primers were designed by the computer program PRIMER v. 0.5 (D. Lincoln & E. Lander, pers. comm.). Criteria for primer pairs were specified by denoting the sequence (target) to be amplified, product size, primer sizes, annealing temperature, and degree of (allowable) complementarity. A printout was produced which listed primer pairs that satisfied these search criteria.

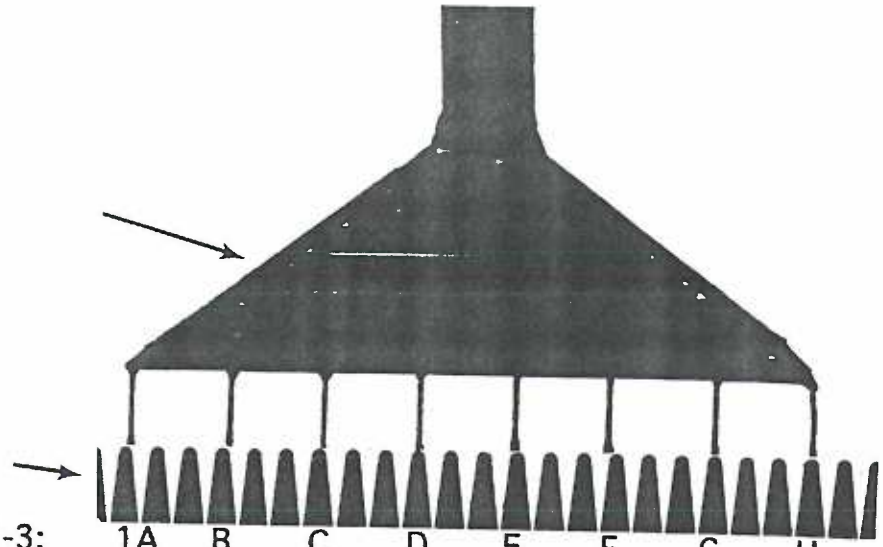
DNA Typing: PCR amplifications of genomic DNA (Saiki *et al.*) were carried out in a total volume of 12.5 µl containing: 25 ng genomic DNA, 10 pmole of each primer, 1.5 mM $MgCl_2$, 200 µM dNTPs, 50 mM KCl, 10 mM Tris-Cl, pH 8.3, 0.6 units Taq polymerase (Perkin Elmer/Cetus) and 0.01% gelatin (Litt 1991a). "Touchdown" amplification (Don *et al.*, 1991) in a Perkin Elmer/Cetus System 9600™ thermocyler was performed. The initial annealing temperature is 10°C above the "target" annealing temperature and is reduced 1°C per cycle for the first 10 cycles. The final 25 cycles are at the target annealing temperature. For all cycles the denaturation was at 94°C for 30 seconds, annealing for 30 seconds and extension at 72°C for 15 seconds. The labeling of amplified product by the incorporation of radioactivity in the PCR, was as above except that unlabeled dCTP is reduced to 20 µM and 0.3 picomoles of [alpha^{32}P]-dCTP is added per sample.

PCR amplification products were resolved on standard sequencing gels (5-6% polyacrylamide, 8M urea in 1X tris-borate buffer) or formamide sequencing gels (6-8% polyarcylamide, 5.6 M urea, 32% formamide in 1X tris-borate buffer). Gels with small or large sharkstooth combs were loaded with 4 ul of sample using single or multichannel pipettors and electrophoresed at 80-100 watts of constant power. Multichannel pipettors allow samples to be loaded onto gels rapidly and in different orientations (Figure 5).

Resolved PCR amplification products were transfered from sequencing gels to Hybond N+ (Amersham) membranes. The membrane was cut to size, prewet in water, placed onto the gel which remained on one glass plate, covered with 3 sheets of dry filter

Figure 5: The gel loading scheme for the multichannel pipettor. The sharkstooth comb forms 96 wells in a sequencing gel. The samples are loaded in rows or columns of 8 from 96 well microtiter plates, each load offset one well from the previous loading forming blocks of 24 samples. The corresponding position on the gel for each microtiter plate well is shown. Long arrow points to the pipettor; the short arrow points to the sharkstooth comb which forms the sample wells in the gel. At the bottom is a schematic of a 96 well microtiter plate.

Figure 5:



Columns 1-3:

1A B C D E F G H

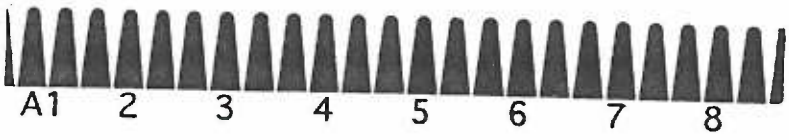
2A B C D E F G H

3A B C D E F G H

Order on gel:

1A2A3A1B2B3B1C2C3C1D2D3D1E2E3E1F2F3F1G2G3G1H2H3H

Rows A-C:



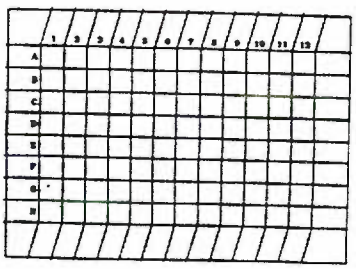
A1 2 3 4 5 6 7 8

B1 2 3 4 5 6 7 8

C1 2 3 4 5 6 7 8

Order on gel:

A1B1C1A2B2C2A3B3C3A4B4C4A5B5C5A6B6C6A7B7C7A8B8C8



paper, a second glass plate and a weight. Transfer of DNA was complete after 2 hours but was usually allowed to proceed overnight. The transfered DNA was fixed to the membrane by submerging it in 0.4 N NaOH for 10 minutes followed with two 5 minute washes in 2xSSC. Alleles were visualized by probing the membrane with locus specific or repeat based probes per previous specifications and autoradiography.

E: Locus Characterization

Mapping microsatellites to a chromosome by two of the following methods:

1) Linkage analysis using CEPH families and the computer program LINKAGE (Lathrop *et al.*, 1984).

2) I*n situ* hybridization of cosmids to human metaphase chromosome spreads.

3) PCR amplification of the Bruns, Mohandas or Patterson somatic cell mapping panels (Bufton *et al.*,1986; Mohandas *et al.*, 1986).

Determination of CA repeat PIC value: Genomic DNA from 80 unrelated CEPH parents were DNA typed for the frequency and number of alleles at a given locus, from which we calculated the % heterozygosity and PIC values (Botstein *et al.*, 1980).

Placing markers on the linkage map: Genotypes for informative CEPH families were determined and Mendelian inheritance confirmed. Initially, families were genotyped until a significant LOD score (>3) could be obtained and linkage demonstrated to a CEPH data base marker using two-point linkage analyses performed by the computer program LINKAGE. Markers which mapped to chromosome 21 were typed by us in all informative CEPH families.

Linkage map construction Theory: The 90-95% confidence limit for the recombination frequency in two point likelihoods is graphically represented by the interval bounded by the

peak LOD score minus 1: [Z(θ) - 1]. (Ott, 1991; Litt, 1987) Typically this interval is broad and spans many centimorgans even when the LOD score is large. For example, the marker D21S11 is tightly linked to the marker D21S13 in CEPH's data base with a maximum LOD score of 12 at theta=0.12; yet the 90-95% confidence interval spans 14 cM (0.06-0.2) For this reason, the order of a set of loci cannot often be determined by two point linkage analysis, since the data can be consistent with several possible orders. Multipoint linkage analysis can be employed in such instances. The following analysis is based on lecture notes by Dr. Mike Litt (Litt 1987). Three point analysis is the simplest case and will be considered here. Given 3 loci there are 3 possible orders: A-B-C; A-C-B; and B-A-C. Any one order will have 4 classes of chromosomes in a fully informative mating. Three of these are recombinant and one is nonrecombinant. If we assume an order of A-B-C, then the classes of chromosomes are defined by recombination occurring in the interval between 1) A&B and A&C; 2) B&C and A&C; 3) A&B and B&C; or 4) no recombination. The likelihoods can be calculated for each class of chromosome and for each possible order of loci. For example, for the order A-B-C, the recombination frequency for each interval is θ_{AB}, θ_{BC} and θ_{AC}. The likelihood for class 1 chromosomes (as shown above) is $\theta_{AB}^{R1}(1-\theta_{BC})^{R1}$ where R1 is the number of class 1 chromosomes. For class 3 the likelihood is $\theta_{AB}^{R3}\,\theta_{BC}^{R3}$ and for nonrecombinant chromosomes $(1-\theta_{AB})^{NR}(1-\theta_{BC})^{N}$ .etc. The likelihood for a particular order is the product of the likelihoods for each class of chromosome from that order. The likelihoods for the other two possible orders are calculated in a similar way. Similar to two point analysis the recombination frequencies for each interval can be varied until a maximum likelihood value is obtained. Likelihood ratios for the 3 possible orders are calculated relative to each other in order to obtain support for one order versus another.

In multipoint analysis the most likely order is that which minimizes the number of double recombinants observed in the data set. This requires that the loci be close enough together such that double recombinants will be infrequent (10-20 cM). For example, the

class 3 chromosome described above should be the least frequent for the order A-B-C. When the data is applied to the other possible orders of loci, the double recombinant chromosome for those orders would presumably not be the least frequent if our original A-B-C order is correct; i.e. only one order should be compatible with the double recombinant chromosome being the least frequent.

The computational complexity of multipoint likelihood analyses are formidable. As the number of loci increase the number of haplotypes and genotypes increase. For example, assuming a fixed order for the 3 loci described above and that each locus has 3 alleles, there are 3x3x3=27 possible haplotypes. The total number of genotypes is 27(27 + 1)/2=378 (Ott, 1991). With five 3 allele loci there are 243 haplotypes and 29,646 genotypes! When considering three loci there are 3 possible orders (mirror images are excluded). As this number increases the possible number of orders increases by n!÷2; therefore with 5 loci there are 60 possible orders. Computer programs have been developed to handle these immense numbers.

Computer programs: We have used two packages of linkage programs to map loci to chromosomes and to construct a linkage map of chromosome 21. One option of LINKAGE (Lathrop *et al.*, 1984) was frequently used to place markers on existing linkage maps- CILINK. In pairwise analysis CILINK iteratively estimates the recombination fraction and calculates the value of theta which gives the maximum likelihood. The linkage program CRIMAP (Green, pers. comm.) was used to construct a map of chromosome 21 with the microsatellite markers developed in this dissertation. Using CRIMAP, two point linkage analysis was performed to identify groups of linked loci. This data was used by BUILD to order loci. First the pair of markers with the greatest number of informative meioses was selected as an ordered pair. Additional markers were inserted sequentially one by one, until all markers were added which met the likelihood tolerence specified prior to running BUILD. Typically the tolerance was set at 1000:1 (LOD=3). Next the FLIPS

option was run in order to examine the likelihoods for order which resulted from the permutation of adjacent loci; this option often detects an incorrectly ordered locus. Finally CHROMPIC was run. This option shows the grandparental origin of each allele and identifies the location of recombinations on each child's chromosome. Double recombination events between adjacent or closely spaced loci identify the most likely data error candidates. The genotypes are rechecked on the original autoradiographs. A genotype is removed from the linkage analysis only if an obvious typing or sample error is detected.

III: RESULTS

Table 1 is summary of the polymorphisms developed in this study. A total of 15 polymorphic microsatellites were characterized; 8 are from chromosome 21. Ten microsatellites have heterozygosities >0.7. Many have been designated reference loci on chromosome index maps under construction for the US genome initiative. The genotypes for 9 polymorphisms were determined for 40 CEPH families and contributed to CEPH. The results for the microsatellites which have been published are detailed in the following manuscripts.

Table 1A: Chromosome 21 microsatellites

Locus	%Het.	Alleles	Position	CEPH#	Reference Marker
D21S11	0.90	14	q21.1	yes	yes
D21S198	0.81	9	q22.3	yes	
D21S167+	0.80	13	q22.2	yes	yes
D21S168+	0.77	8	q22.3	yes	
D21S172	0.71rev	8	q11.2	yes	
D21S236	0.71	7	q11.2	yes	
D21S218	0.55	4	q22.1	yes	
D21S13+	0.42	5	q11.2	no	

Table 1B: Non Chromosome 21 microsatellites

Locus	%Het.	Alleles	Position	CEPH#	Reference Marker*
D11S35+	0.79	6	11q22	no	yes
D1S117	0.77	11	1q23-q25	no	
D9S55	0.75	9	9p12	no	yes
D14S43	0.72	10	14q24.3	yes	yes
D4S233	0.69	6	4q	no	
D18S37	0.65	6	18p	yes	yes^
D1S116	0.62	7	1p31.2	no	

\# Markers whose genotypes have been contributed to CEPH.
* As indicated in Cytogenet. Cell Genet./HGM11
^ Richard Straub pers. comm.
\+ VS second author

A: Chromosome 21 Manuscripts

1: D21S11- Human Molecular Genetics (1992), Vol 1, No. 1, pg. 67

Tetranucleotide repeat polymorphism at the D21S11 locus

V.Sharma and M.Litt
Departments of Biochemistry and Molecular Biology and Molecular and Medical Genetics, Oregon Health Sciences University, Portland, OR 97201, USA

Source and Description: Cosmid ICRFc102B06134 was from a flow sorted chromosome 21 cosmid library. DNA sequences flanking a $(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3TA(TCTA)_3TC(AT-CT)_{10}$ repeat within an Alu I subclone (VS17TB2) of this cosmid (GenBank accession no. M84567) were used to design PCR primers.

PCR Primers:
VS17T #3: 5'-GTGAGTCAATTCCCCAAG-3'
VS17T #4: 5'-GTTGTATTAGTCAATGTTCTCC-3'

Polymorphism: Allelic fragments were resolved on 4.5% denaturing acrylamide gels.

Allele	Length(nt)	Frequency	Allele	Length(nt)	Frequency
E1	264	.006	E7	224	.032
E2	248	.013	E8	220	.247
E3	240	.117	E9	212	.221
E4	236	.026	E10	204	.156
E5	232	.065	E11	196	.013
E6	228	.097	E12	172	.006

Alleles found in four CEPH parents were: 133101 = A3, A9; 133102 = A4, A5; 137701 = A7, A8; 137702 = A9, A11. Allele frequencies were measured in 72 unrelated European Caucasians. The heterozygosity was 90% and the PIC was 0.82.

Chromosomal Localization and Mendelian Inheritance: Cosmid ICRFc102B06134 was identified by hybridization with the D21S11 probe pPW236B. Linkage analysis with the D21S11 EcoRI RFLP in 5 CEPH families (50 informative meioses) gave a maximum LOD score of 14.1 at $\theta = 0$. Mendelian inheritance was observed in all cases.

PCR Conditions: PCR was performed in a total volume of 12.5 μl using standard buffer and dNTP concentrations according to (1) except that 0.25 mM spermidine was included and the products were not radioactively labelled. 'Touchdown' amplification (2) in a Perkin-Elmer/Cetus System 9600™ thermocycler was performed, using an initial annealing temperature of 62° for 1 min which was decreased by 1 degree per cycle for the first 10 cycles. For the final 25 cycles, the annealing temperature was held constant at 52° and the annealing time was 30 sec. For all cycles, denaturation was for 1 min at 94° and extension was for 1 min at 72°. PCR products were resolved on DNA sequencing gels, capillary-blotted onto Hybond N+™ membranes and detected by probing with 5'[^{32}P] labeled $(GATA)_7$ oligomer.

References: 1) Lutty,J.A., Guo,Z., Willard,H.F., Ledbetter,D.H., Ledbetter,S. and Litt,M. (1990) *Am. J. Hum. Genet.* 46, 776–783. 2) Don,R.H., Cox,P.T., Wainwright,B.J., Baker,K. and Mattick,J.S. (1991) *Nucl Acids Res.* 19, 4008.

2: D21S236- Human Molecular Genetics (1992), Vol. 1, No. 4, pg 289

Dinucleotide repeat polymorphism at the D21S236 locus

V.Sharma, P.de Jong[1], K.Yokobata[1] and M.Litt*
Departments of Biochemistry and Medical Genetics, Oregon Health Sciences University, Portland, OR 97201 and [1]Biomedical Sciences Division, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

Source and Description: Cosmid 21A5-J52−0 was from a flow sorted chromosome 21 cosmid library LL21NCO1. DNA sequences flanking a $(TG)_{11}$ repeat within an AluI subclone (JA5) of this cosmid (GenBank accession no. M88274) were used to design PCR primers.

PCR Primers:

JA5#3: 5′-CCCAAATAAAAAAGAGAACAG-3′
JA5#4: 5′-CTAAAGAGGACTTCAGAGTAAGG-3′

Polymorphism: Allelic fragments were resolved on DNA sequencing gels.

Allele	Length (nt)	Frequency
A1	128	.04
A2	126	.08
A3	124	.24
A4	122	.14
A5	120	.02
A6	106	.01
A7	104	.47

Alleles found in four CEPH parents were: 133201 = A2,A7; 133202 = A3,A4; 134501 = A1,A2; 134502 = A2,A7. Allele frequencies were measured in 76 unrelated CEPH parents. Heterozygosity was 71% and the PIC calculated from these frequencies was 0.65.

Chromosomal Localization and Mendelian Inheritance: Fluorescent *in situ* hybridization of cosmid DNA to human chromosome metaphase spreads indicated localization to band q11.2. Using the CEPH database v5, linkage analysis with 12 CEPH families gave a maximum LOD score of 14.1 at theta = 0.028 with D21S13 (probe D21K9, Taq I RFLP) in 106 informative meioses. Mendelian inheritance was observed in all cases.

PCR Conditions: PCR reactions were carried out in a total volume of 12.5 μl containing: 25 ng genomic DNA, 15 pmole of each primer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 50 mM KCl, 10 mM tris-Cl, pH 8.3 and 0.375 units of Taq polymerase (Ampli-Taq). Two temperature 'touchdown' PCR amplification (1) in a Perkin-Elmer/Cetus System 9600 thermocycler was performed, using an initial annealing/extension temperature of 61°C for 30 seconds which was decreased by one degree per cycle for the first ten cycles. For the final 20 cycles, the annealing/extension temperature was held constant at 51°C for 30 seconds. For all cycles, denaturation was at 94°C for 10 seconds. PCR products were resolved on 7% acrylamide gels containing 5.6 M urea plus 32% formamide, capillary-blotted onto Hybond N^+ membrane and detected by probing with 5′[^{32}P] labeled $(CA)_{15}$ oligomer.

Acknowledgements: This work was supported by NIH grant HG00022 and DOE contract No. W-7405-Eng-48.

Reference: 1) Don,R.H., Cox,P.T., Wainwright,B.J., Baker,K. and Mattick,J.S. (1991) *Nucleic Acids Res.* **19**, 4008.

* To whom correspondence should be addressed

3: D21S172- Nucleic Acids Research (1991), Vol. 19, No. 5 pg. 1169

Dinucleotide repeat polymorphism at the D21S172 locus

V.Sharma, L.Allen, R.E.Magenis and M.Litt*
Department of Medical Genetics, Oregon Health Sciences University, Portland, OR 97201, USA

Source and Description: Cosmid ICRFc102H0921 was from a flow sorted chromosome 21 cosmid library. DNA sequence flanking a $(CA)_{20}$ repeat within a Sau3A subclone (1H9) of this cosmid (EMBL accession # X56513) were used to design PCR primers.

PCR Primers:
1H9#2: 5'-TACAGTGGCAAATGTCATTG-3'
1H9#3: 5'-GAATATGTGTTAGGTCCTGC-3'

Polymorphism: Eight allelic fragments were resolved on DNA sequencing gels. Lengths of the allelic fragments (nt) were: A1 = 161, A2 = 157, A3 = 155, A4 = 153, A5 = 151, A6 = 149, A7 = 147, A8 = 145.

Frequencies: Allele frequencies in 51 unrelated European Caucasians were:
A1 = .01
A2 = .029
A3 = .098
A4 = .049
A5 = .588
A6 = .176
A7 = .02
A8 = .029
PIC = .58

Chromosomal Localization and Mendelian Inheritance: Fluorescence in situ hybridization localized cosmid ICRFc102H0921 to 21q11.2. Linkage with 4 CEPH families showed no recombination in 35 informative meioses with D21S13 probe D21K9 (localized to q11.2), giving a maximum LOD score of 9.62 at theta = 0. Mendelian inheritance was observed in all cases.

PCR Conditions: PCR reactions are carried out in a total volume of 25 μl containing: 25 ng genomic DNA, 10 pmole of each primer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 50 mM KCl, 10 mM tris-Cl, pH 8.3, 0.6 units Taq polymerase (Cetus) and 0.01% gelatin. Amplification is for 35 cycles with denaturation at 94°C for 45 seconds, annealing at 56°C for 45 seconds and extension at 72°C for 45 seconds.

Acknowledgements: This work was supported by NIH grant HG00022. V. Sharma is the recipient of a fellowship from the Medical Research Foundation of Oregon. We thank H. Lehrach for providing us with ICRFc102H0921.

4: D21S198- Nucleic Acids Research (1991), Vol. 19, No. 14, pg 4023

Dinucleotide repeat polymorphism at the D21S198 locus

V.Sharma[1], D.Patterson[2] and M.Litt[1]
[1]Department of Medical Genetics, Oregon Health Sciences University, Portland, OR 97201 and [2]Eleanor Roosevelt Institute for Cancer Research, Denver, CO 80206, USA

Source and Description: Cosmid ICRFc102D1122 was from a flow sorted chromosome 21 cosmid library. DNA sequences flanking a $(TG)_{23}$ repeat within a HaeIII subclone (VS1) of this cosmid (EMBL accession no. X58124) were used to design PCR primers.

PCR primers:
VS1#1: 5'-TGTCACATAAGGGAAGAG-3'
VS1#2: 5'-GATGTTAGCAGCTCTGC-3'

Polymorphism: Nine allelic fragments were resolved on DNA sequencing gels. Lengths of the allelic fragments (nt) were: A1 = 128, A2 = 126, A3 = 124, A4 = 122, A5 = 120, A6 = 118, A7 = 116, A8 = 114, A9 = 112. Alleles found in four CEPH parents were: 140801 = A1, A4; 140802 = A6, A8; 142301 = A1, A6; 142302 = A5, A9.

Frequencies: Allele frequencies in 44 unrelated European Caucasians were:
A1 = .045 A2 = .068 A3 = .068 A4 = .023
A5 = .113 A6 = .205 A7 = .068 A8 = .295
A9 = .113 PIC = .81

Chromosomal Localization and Mendelian Inheritance: PCR of genomic DNAs from a somatic cell hybrid panel (1) indicated localization to 21q22.3. Linkage analysis with 3 CEPH families showed no recombination in 24 informative meioses with PFKL probe pG-PFKL3.3 (localized to q22.3 (2)), giving a maximum LOD score of 6.4 at theta = 0. Mendelian inheritance was observed in all cases.

PCR Conditions: PCR reactions are carried out in a total volume of 25 μl containing: 25 ng genomic DNA, 15 pmole of each primer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 50 mM Tris-Cl, 20 mM $(NH_4)_2SO_4$, pH 9 and 0.6 units of Replinase™ (NEN/Dupont). Amplification is for 35 cycles with denaturation at 95°C for 60 seconds, annealing at 54°C for 60 seconds and extension at 72°C for 30 seconds. PCR products were resolved on DNA sequencing gels, capillary-blotted onto Hybond N+™ nylon membranes (Amersham) and revealed by probing with 5'[^{32}P] labeled $(TG)_7$ oligomer.

References: 1) K.Gardiner, M.Horisberger, J.Kraus, U.Tantravahi, J.Korenberg, V.Rao, S.Reddy and D.Patterson (1990) *EMBO J.* 9, 25–34. 2) B.Carritt and M.Litt (1989) *Cytogenet Cell Genet* 51, 358–371.

5: D21S218- Human Molecular Genetics (1992), submitted

Dinucleotide repeat polymorphism at the D21S218 locus.

V. Sharma*, P. de Jong#, K. Yokobata#, and M. Litt*, *Departments of Biochemistry and Medical Genetics, Oregon Health Sciences University, Portland, OR 97201 USA. # Biomedical Sciences Division, Lawrence Livermore National Laboratory, Livermore, CA 94550.

Source and description : Cosmid 21B2 J52-0 was from a flow-sorted human chromosome 21 specific cosmid library **LL21NCO1**. DNA sequences flanking a $(TG)_{13}$ repeat within Alu I subclone VJB2 of this cosmid (EMBL accession # 62652) were used to design PCR primers.

PCR primers: JB2#1 = 5'-GGTCTGCCATTCTTACCTCACAAGA-3'
JB2#2 = 5'-ACCATTTAAATGCATCTCCAGATTCG-3'

Polymorphism: Allelic fragments were resolved on DNA sequencing gels.

Allele	Length(nt)	Frequency
A1	124	.06
A2	122	.09
A3	120	.58
A4	118	.27

Alleles in four CEPH parents were as follows: 133101: A2,A4; 133102: A3,A4; 133201: A1,A3; 133202: A3, A3. Allele frequencies were measured in 44 unrelated CEPH parents. The PIC calculated from these frequencies is 0.52.

Chromsomal localization and Mendelian inheritance: Fluorescent in situ hybridization of cosmid DNA to human metaphase chromosome spreads indicated localization to 21q22.1. Linkage analysis with D21S82 in 9 informative CEPH families gave a maximum LOD score of 10.5 at theta= 0.041. Mendelian inheritance was observed in all cases.

PCR conditions: PCR was performed in a total volume of 12.5 µl using standard buffer and dNTP concentrations according to (1) except that the products were not radioactively labelled. Amplification in a Perkin-Elmer/Cetus System 9600™ thermocycler was for 30 cycles with denaturation at 94° for 15 sec, annealing at 63° for 30 sec and extension at 72° for 15 sec. PCR products were resolved on DNA sequencing gels, capillary-blotted onto Hybond N+ ™ nylon membranes (Amersham) and revealed by probing with 5'[^{32}P] labeled $(CA)_{15}$ oligomer.

Reference: 1) Luty, J.A., Guo, Z., Willard, H.F., Ledbetter, D.H., Ledbetter, S. and Litt , M.(1990) Am. J . Hum. Genet. 46:776-783.

Acknowledgements: This work was supported by NIH Grant HG00022 and DOE contract No. W-7405-Eng-48.

6: D21S167- Nucleic Acids Research (1990), Vol. 18, No. 16, pg. 4967

TG repeat polymorphism at the D21S167 locus

Z.Guo, V.Sharma, D.Patterson[1] and M.Litt*
Oregon Health Sciences University, Portland, OR 97201 and [1]Eleanor Roosevelt Institute for Cancer Research, Denver, CO 80206, USA

Source and Description: Cosmid ICRFc102A0322 was from a flow-sorted human chromosome 21-specific library. DNA sequences flanking a $(GT)_{19}$ repeat within this cosmid (EMBL accession number X52289) were used to design PCR primers.

PCR Primers:

(# 112)-5'-TCCTTCCATGTACTCTGCA-3'
(# 113)-5'-TGCCCTGAAGCACATGTGT-3'

Polymorphism: Allelic fragments were resolved on DNA sequencing gels. Lengths of allelic fragments (nt) were: A1 = 156, A2 = 158, A3 = 160, A4 = 162, A5 = 164, A6 = 166, A8 = 172, A9 = 174, A10 = 176, A11 = 178, A12 = 180, A13 = 182.

Frequencies: Allele frequencies in 33 unrelated European Caucasians: A1 = .030, A2 = .272, A3 = 288, A4 = .061, A5 = .030, A6 = .030, A7 = .061, A8 = .061, A9 = .030, A10 = .061, A11 = .045, A12 = .015, A13 = .015; PIC = 0.80.

Chromosomal Localization and Mendelian Inheritance: PCR of genomic DNAs from a somatic cell hybrid panel (1) indicated localization to 21q22.2. Linkage analysis in 3 CEPH families showed no recombination in 21 informative phase-known meioses with D21S15 (previously localized to 21q22.2 (1)), giving a maximum LOD score of 6.3 at theta = 0. Mendelian inheritance was observed in all cases.

PCR Conditions: We carry out PCR as in reference (2) using an annealing temperature of 56°C.

Acknowledgements: This work was supported by Grants GM-32500 and HD17449 from the National Institutes of Health. We thank H.Lehrach for providing us with ICRFc102A0322.

References: 1) Gardiner,K., Horisberger,M., Kraus,J., Tantravahi,U., Korenberg,J., Rao,V., Reddy,S. and Patterson,D. (1990) *EMBO J.* **9**, 25–34. 2) Luty,J.A., Guo,Z., Willard,H.F., Ledbetter,D.H., Ledbetter,S. and Litt,M. (1990) *Am. J. Hum. Genet.* **46**, 776–783.

7: D21S168- Nucleic Acids Research (1990), Vol. 18, No. , pg. 5924

Dinucleotide repeat polymorphism at the D21S168 locus

Z.Guo, V.Sharma, D.Patterson[1] and M.Litt*
Oregon Health Sciences University, Portland, OR 97201
and [1]Eleanor Roosevelt Institute for Cancer Research,
Denver, CO 80206, USA

Source and Description: Cosmid ICRFc102D0494 was from a flow-sorted human chromosome 21-specific library. DNA sequences flanking a $(GT)_{19}$ repeat within AluI subclone C21G-94A of this cosmid (EMBL accession number X53367) were used to design PCR primers.

PCR Primers: (#334)-5'-ATGCAATGTTATGTAGGCTG-3'
(#335)-5'-CGGCATCACAGTCTGATAAA-3'

Polymorphism: Allelic fragments were resolved on DNA sequencing gels. Lengths of allelic fragments (nt) were: A1=118, A2=116, A3=114, A4=112, A5=110, A6=108, A7=106, A8=104.

Frequencies: Allele frequencies in 32 unrelated European Caucasians: A1=.047, A2=.063, A3=.25, A4=.34, A5=.22, A6=.047, A7=.016, A8=.016, PIC=0.73.

Chromosomal Localization and Mendelian Inheritance: PCR of genomic DNAs from a somatic cell hybrid panel (1) indicated localization to Region C1, corresponding to 21q22.3. Linkage analysis in 2 CEPH families showed no recombination in 14 informative meioses with D21S3 (previously localized to 21q22.3 (1)), giving a maximum LOD score of 3.0 at theta=0. Mendelian inheritance was observed in all cases.

PCR Conditions: We carry out PCR in a total volume of 25 μl containing: 50 ng genomic DNA, 5 pmoles of each primer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 50 mM KCl, 10 mM Tris-Cl^-, pH 8.3, 0.6 units Taq polymerase (Perkin-Elmer/Cetus) and 0.01% gelatin. Amplification is for 30 cycles with denaturation at 94°C for 1 min, annealing at 53°C for 1 min and extension at 72°C for 30 sec.

Acknowledgements: This work was supported by Grants GM-32500 and HD17449 from the National Institutes of Health. We thank H.Lehrach for providing us with ICRFc102D0494.

References: 1) Gardiner,K., Horisberger,M., Kraus,J., Tantravahi,U., Korenberg,J., Rao,V., Reddy,S. and Patterson,D. (1990) *EMBO J.* 9, 25–34. 2) Luty,J.A., Guo,Z., Willard,H.F., Ledbetter,D.H., Ledbetter,S. and Litt,M. (1990) *Am. J. Hum. Genet.* 46, 776–783.

8: D21S13E- Nucleic Acids Research (1990), Vol. 18, No. 24, pg. 7470

Dinucleotide repeat polymorphism at the D21S13E locus

Z.Guo, V.Sharma and M.Litt*
Oregon Health Sciences University, Portland, OR 97201, USA

Source/Description: Cosmid ICRFc102B11128 was from a flow-sorted human chromosome 21-specific library. DNA sequences flanking an $(AC)_{18}$ repeat within AluI subclone C21G-K23A of this cosmid (EMBL accession number X54391) were used to design PCR primers.

PCR Primers: (50)-5'-ATCTGATGCCTTTGAATCT-3'
(51)-5'-TATGCCATTCACTGTGATA-3'

Polymorphism: Allelic fragments were resolved on DNA sequencing gels. Lengths of allelic fragments (nt) were: A1 = 125, A2 = 123, A3 = 121, A4 = 117, A5 = 111.

Frequencies: Allele frequencies in 62 unrelated CEPH parents: F1 = .024, F2 = .024, F3 = .69, F4 = .024, F5 = .23. PIC = 0.42. This system is in linkage equilibrium with the TaqI RFLP revealed by probe D21K9 (1) (χ^2 = 2.73, df = 4, $p>0.5$). With the combined systems, 9 haplotypes were observed in 55 unrelated CEPH parents and the combined PIC was 0.69.

Chromosomal Localization and Mendelian Inheritance: Cosmid ICRFc102B11128 was identified by screening with the D21S13E clone D21K9. Linkage analysis in 8 CEPH families showed no recombination in 46 informative meioses with the TaqI RFLP identified by D21K9 (1) (previously localized to 21q11.2 (2)), giving a maximum LOD score of 13.2 at theta = 0. Mendelian inheritance was observed in all cases.

PCR Conditions: We carry out PCR in a total volume of 25 μl containing: 25 ng genomic DNA, 5 pmoles of each primer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 50 mM KCl, 10 mM Tris-Cl^-, pH 8.3, 0.6 units Taq polymerase (Perkin-Elmer/Cetus) and 0.01% gelatin. Amplification is for 30 cycles with denaturation at 94°C for 1 min, annealing at 54°C for 1 min and extension at 72°C for 30 sec.

References: 1)Davies,K.E., Harper,K., Bonthron,D., Krumlauf,R., Polkey,A., Pembrey,M.E. and Williamson,R. (1984) *Hum. Genet.* 66, 54–56. 2) Gardiner,K., Horisberger,M., Kraus,J., Tantravahi,U., Korenberg,J., Rao,V., Reddy,S. and Patterson,D. (1990) *EMBO J.* 9, 25–34.

Acknowledgements: This work was supported by Grants GM-32500 and HG-00022 from the National Institutes of Health. We thank H.Lehrach for providing us with ICRFc102B11128.

B: Non Chromosome 21 Repeats

1. D18S37- Human Molecular Genetics (1992), Vol. 1, No. 4, pg. 289

Dinucleotide repeat polymorphism at the D18S37 locus

V.Sharma, Z.Guo and M.Litt*
Departments of Biochemistry and Medical Genetics, Oregon Health Sciences University, Portland, OR 97201, USA

Source and Description: Cosmid ICRFc102H1021 was from a flow sorted chromosome 21 cosmid library. DNA sequences flanking a $(AC)_{15}$ repeat within HaeIII subclone C21G12 of this cosmid (GenBank accession no. M88273) were used to design PCR primers.

PCR Primers:

#21: 5′-ACTGCAACAATGCTGCAAT-3′
#22: 5′-AAATGAGAGAATCAGTGGC-3′

Polymorphism: Allelic fragments were resolved on DNA sequencing gels.

Allele	Length (nt)	Frequency
A1	108	.01
A2	106	.125
A3	104	.485
A4	102	.36
A5	100	.01
A6	98	.01

Alleles found in four CEPH parents were: 1329101 = A3, A4; 1329102 = A1, A3; 134401 = A2, A4; 134402 = A3, A3. Allele frequencies were measured in 76 unrelated CEPH parents. Heterozygosity was 65% and the PIC calculated from these frequencies was 0.55.

Chromosomal Localization and Mendelian Inheritance: PCR of genomic DNAs from somatic cell hybrid panels (1, 2) indicated localization to chromosome 18. Linkage analysis with 16 CEPH families using data from the CEPH database v5 gave a maximum LOD score of 12.4 at theta = 0.02 with D18S12 (CRI-R397/PstI) in 145 informative meioses. Mendelian inheritance was observed in all cases.

PCR Conditions: PCR reactions were carried out in a total volume of 12.5 μl containing: 25 ng genomic DNA, 15 pmole of each primer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 50 mM KCl, 10 mM Tris−Cl, pH 8.3, and 0.6 units of Taq polymerase (Ampli-Taq). Amplification in a Perkin-Elmer/Cetus System 9600 thermocycler was for 30 cycles with denaturation at 94°C for 15 seconds, annealing at 52°C for 30 seconds and extension at 72°C for 15 seconds. PCR products were resolved on 7% acrylamide gels containing 5.6 M urea plus 32% formamide, capillary-blotted onto Hybond N^+ membrane and detected by probing with 5′[^{32}P] labeled $(CA)_{15}$ oligomer.

References: 1) Bufton,L., Bruns,G.A., Magenis,R.E., Tomar,D., Shaw,D. and Litt,M. (1986) *Am. J. Hum. Genet.* 38, 447−460. 2) Mohandas,T., Heinzmann,C., Sparkes,R.F., Wasmuth,J., Edwards,P. and Lusis,A. (1986) *Somatic Cell. Molec. Genet.* 12, 89−94.

* To whom correspondence should be addressed

2. D14S43- Nucleic Acids Research (1991), Vol. 19, No. 7, pg. 1722

Dinucleotide repeat polymorphism at the D14S43 locus

V.Sharma, L.Smith, L.Allen, R.E.Magenis and M.Litt*
Department of Medical Genetics, Oregon Health Sciences University, Portland, OR 97201, USA

Source/Description: Cosmid ICRFc102E1222 was from a flow sorted chromosome 21 cosmid library. DNA sequences flanking the sequence $(TG)_{14}$TTGGGAATGGGGCTGGGGGGAGAC-AGGGT$(TG)_{10}$ within HaeIII subclone 2E12B of this cosmid (EMBL accession number X56973) were used to design PCR primers.

PCR Primers: 2E12B#1: 5'-TGGAACACTCAGGCGA- 3'
2E12B#2: 5'-CCAGAGCCACTTTCTAC-3'

Polymorphism: Ten allelic fragments were resolved on DNA sequencing gels. Lengths of allelic fragments (nt) were: A1 = 193, A2 = 191, A3 = 189, A4 = 187, A5 = 185, A6 = 183, A7 = 181, A8 = 161, A9 = 159, A10 = 155.

Frequencies: Allele frequencies in 52 unrelated Caucasians were: A1 = .01, A2 = .01, A3 = .11, A4 = .13, A5 = .26, A6 = .02, A7 = .06, A8 = .01, A9 = .38, A10 = .01; PIC = 0.72.

Chromosomal Localization and Mendelian Inheritance: PCR of genomic DNAs from two somatic cell hybrid panels indicated localization to chromosome 14. Fluorescent in situ hybridization indicated localization to 14q24.3. Linkage analysis in 8 CEPH families with 85 informative meioses against locus D14S24 (probe CRI-C70) gave a maximum LOD score of 2.95 at $\theta = 0.3$. Mendelian inheritance was observed in all cases.

PCR Conditions: PCR reactions are carried out in a total volume of 25 μl containing: 25 ng genomic DNA, 10 pmole of each primer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 50 mM KCl, 10 mM Tris−Cl, pH 8.3, 0.6 units Taq polymerase (Ampli-Taq) and 0.01% gelatin. Amplification is for 35 cycles with denaturation at 94°C for 60 seconds, annealing at 54°C for 60 seconds and extension at 72°C for 30 seconds.

Comments: This marker locus will be a useful index marker (1) for the genetic linkage map of chromosome 14.

Acknowledgements: This work was supported by NIH Grant HG00022. V.Sharma is the recipient of a fellowship from the Medical Research Foundation of Oregon. We thank H.Lehrach for providing cosmid ICRFc102E1222.

Reference: 1) Roberts,L. (1990) *Science* 248, 805.

3. **D9S56-** Nucleic Acids Research (1991), Vol. 19, No. 14, pg 4023

Dinucleotide repeat polymorphism at the D9S55 locus

V.Sharma, D.Browne,L.Smith, R.E.Magenis and M.Litt
Department of Medical Genetics, Oregon Health Sciences University, Portland, OR 97201, USA

Source and Description: Cosmid ICRFc102F0621 was from a flow sorted chromosome 21 cosmid library. DNA sequences flanking a $(TG)_{24}$ repeat within an EcoRI subclone (1F6) of this cosmid (EMBL accession no. X58125) were used to design PCR primers.

PCR primers:
1F6#2: 5'-GAGAAAATTCCAGGCA-3'
1F6#3: 5'-GGTTGAGTCGTTCTTA-3'

Polymorphism: Nine allelic fragments were resolved on DNA sequencing gels. Lengths of the allelic fragments (nt) were: A1 = 197, A2 = 191, A3 = 189, A4 = 187, A5 = 185, A6 = 183, A7 = 181, A8 = 179, A9 = 165. Alleles found in four CEPH parents were: 134901 = A2, A3; 134902 = A5, A7; 141601 = A6, A7; 141602 = A3, A4.

Frequencies: Allele frequencies in 57 unrelated European Caucasians were:

A1 = .01	A2 = .01
A3 = .170	A4 = .315
A5 = .245	A6 = .130
A7 = .105	A8 = .01
A9 = .01	
PIC = .75	

Chromosomal Localization and Mendelian Inheritance: Fluorescence in situ hybridization localized cosmid ICRFc102F0621 to 9p12. Linkage analysis with 4 CEPH families showed no recombination in 40 informative meioses with D9S18 probe pHHH220 (localized to 9p13 (1)), giving a maximum LOD score of 11.5 at theta = 0. Mendelian inheritance was observed in all cases.

PCR Conditions: PCR reactions are carried out in a total volume of 25 μl containing: 25 ng genomic DNA, 15 pmole of each primer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 50 mM tris-Cl, 20 mM $(NH_4)_2SO_4$, pH 9 and 0.6 units of Replinase™ (NEN/Dupont). Amplification is for 35 cycles with denaturation at 95°C for 60 seconds, annealing at 52°C for 60 seconds and extension at 72°C for 30 seconds. PCR products were resolved on DNA sequencing gels, capillary-blotted onto Hybond N+™ nylon membranes (Amersham) and revealed by probing with 5'[^{32}P] labeled $(TG)_7$ oligomer.

Reference: 1) M.Lathrop, Y.Nakamura, P.O'Connell, M.Leppert, S.Woodward, J.M.Lalouel and R.White (1988) *Genomics* 3, 361–366.

4. D1S117- Nucleic Acids Research (1991), Vol 19, No. 5, pg. 1168

Dinucleotide repeat polymorphism at the D1S117 locus

V.Sharma and M.Litt*
Department of Medical Genetics, Oregon Health Sciences University, Portland, OR 97201, USA

Source and Description: Cosmid ICRFc102H1222 was from a flow sorted chromosome 21 cosmid library. DNA sequences flanking the sequence $(TG)_{21}(TA)_{10}$ within a HaeIII subclone (2H12A) of this cosmid (EMBL accession # X54558) were used to design PCR primers.

PCR Primers:
2H12A # 1: 5'-CCTTTTGCCTCCTTCGT-3'
2H12A # 2: 5'-CTCATTTACAATAGCTACC-3'.

Polymorphism: Eleven allelic fragments were resolved on DNA sequencing gels. Lengths of allelic fragments (nt) were: A1 = 132, A2 = 126, A3 = 124, A4 = 122, A5 = 120, A6 = 118, A7 = 116, A8 = 114, A9 = 112, A10 = 110, A11 = 100.

Frequencies: Allele frequencies in 28 unrelated European Caucasians were:

A1	= .018	A7	= .018
A2	= .036	A8	= .036
A3	= .054	A9	= .071
A4	= .125	A10	= .036
A5	= .357	A11	= .036
A6	= .214	PIC	= 0.77

Chromosomal Localization and Mendelian Inheritance: PCR of genomic DNAs from a somatic cell hybrid panel indicated localization to chromosome 1. Linkage analysis in 4 CEPH families with 28 informative meioses against locus AT3 (localized to q23−q25.1 (1)), gave a maximum LOD score of 5.06 at theta = 0.065. Mendelian inheritance was observed in all cases.

PCR Conditions: PCR reactions are carried out in total volume of 25 μl containing: 25 ng genomic DNA, 10 pmole of each primer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 50 mM KCl, 10 mM tris-Cl, pH 8.3, 0.6 units Taq polymerase (BRL) and 0.01% gelatin. Amplification is for 35 cycles with denaturation at 94°C for 60 seconds, annealing at 46°C for 60 seconds and extension at 72°C for 30 seconds.

Acknowledgements: This work was supported by grant HG00022 from the National Institutes of Health. V. Sharma is the recipient of a fellowship from the Medical Research Foundation of Oregon. We thank H. Lehrach for providing cosmid ICRFc102H1222.

Reference: 1) Bruns,G.A.P. and Sherman.S.L. (1989) *Cytogenet. Cell Genet.* 51, 67−90.

5. D1S116-

Nucleic Acids Research (1991), Vol. 19, No. 5, pg. 1169

A dinucleotide repeat polymorphism at the D1S116 locus

V.Sharma, L.Allen, R.E.Magenis and M.Litt*
Oregon Health Sciences University, Portland, OR 97201, USA

Source and Description: Cosmid ICRFc102D0922 was from a flow sorted chromosome 21 cosmid library. DNA sequences flanking a $(GT)_{17}$ repeat within an AluI subclone (2D9-1) of this cosmid (EMBL accession # X54557) were used to design PCR primers.

PCR Primers:
2D9#1–5'-TACAAGGCAACCACATAATT-3'
2D9#2–5'-CTTTTCCTAATTGTGTGTGT-3'

Polymorphism: Allelic fragments were resolved on DNA sequencing gels. Lengths of allelic fragments (nt) were: A1 = 101, A2 = 99, A3 = 97, A4 = 95, A5 = 93, A6 = 91, A7 = 89.

Frequencies: Allele frequencies in 39 unrelated European Caucasians were:
A1 = .077
A2 = .103
A3 = .217
A4 = .013
A5 = .013
A6 = .051
A7 = .525
PIC = 0.62

Chromosomal Localization and Mendelian Inheritance: Fluorescence in situ hybridization localized cosmid ICRFc102D0922 to 1p31.2. Linkage analysis in 2 CEPH families showed no recombination in 14 informative meioses with D1S22 (localized to pter-p22 (1)), giving a maximum LOD score of 4.50 at theta = 0. Mendelian inheritance was observed in all cases.

PCR Conditions: We carry out PCR in a total volume of 25 μl containing: 25 ng genomic DNA, 10 pmole of each primer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 50 mM KCl, 10 mM tris-Cl, pH 8.3, 0.6 units Taq polymerase (BRL) and 0.01% gelatin. Amplification is for 35 cycles with denaturation at 94°C for 30 seconds, annealing at 46°C for 45 seconds and extension at 72°C for 1.5 minutes.

Acknowledgements: This work was supported by grant HG00022 from the National Institutes of Health. V. Sharma is the recipient of a fellowship from the Medical Research Foundation of Oregon. We thank H. Lehrach for providing cosmid ICRFc102D0922.

Reference: 1) Bruns,G.A.P. and Sherman,S.L. (1989) *Cytogenet. Cell Genet.* 51, 67–90.

* To whom correspondence should be addressed

6. D4S233- Human Molecular Genetics (1992), submitted

Dinucleotide Repeat Polymorphism at the D4S233 Locus. V. Sharma and M. Litt*, Departments of Biochemistry and Molecular Biology and Molecular and Medical Genetics, Oregon Health Sciences University, Portland, OR 97201 USA.

Source and Description: Cosmid ICRFc102A07119 was from a flow sorted chromosome 21 cosmid library. DNA sequences flanking a $(TG)_{13}$ repeat within an Alu I subclone (VS17) of this cosmid (EMBL accession # Z11164) were used to design PCR primers.

PCR primers: VS17#1: 5'-CATACCACTGCACTCCAGTC-3'
VS17#2: 5'-GCATATCAAATAGTTGGGCAAG-3'

Polymorphism: Allelic fragments were resolved on DNA sequencing gels.

Allele	Length(nt)	Frequency
A1	151	.04
A2	145	.025
A3	139	.21
A4	137	.105
A5	133	.21
A6	131	.41

Alleles in four CEPH parents were as follows: 137501: A1,A6; 137502: A3,A6; 142101: A3,A4; 142102: A2, A5. Allele frequencies were measured in 76 unrelated CEPH parents. The PIC calculated from these allele frequencies was 0.69.

Chromosomal Localization and Mendelian Inheritance: PCR of genomic DNAs from somatic cell hybrid panels (1,2) indicated localization to chromosome 4. Linkage analysis with 5 CEPH families gave a maximum LOD score of 7.9 at theta=0.025 with FGB probe FIB Bcl IB in 41 informative meioses. Mendelian inheritance was observed in all cases.

PCR Conditions: PCR reactions are carried out in a total volume of 12.5 µl containing: 25 ng genomic DNA, 20 pmole of each primer, 1.5 mM $MgCl_2$, 200 uM dNTPs, 50 mM KCl, 10 mM tris-Cl, pH 8.3, and 0.6 units of Taq polymerase (Ampli-Taq). Amplification in a Perkin- Elmer/Cetus System 9600™ thermocycler was for 30 cycles with denaturation at 94° for 10 seconds and annealing/extension at 62.5° for 30 seconds.

References: 1) Bufton,L., Bruns, G.A.., Magenis, R.E.,Tomar, D., Shaw,D., Brook, D., and Litt,M. (1986) Am. J. Hum. Genet. 38, 447-460. 2.)Mohandas,T., Heinzmann, C., Sparkes, R.F., Wasmuth, J., Edwards, P. and Lusis, A.(1986) Somatic Cell. Molec. Genet. 12:89-94.

Acknowledgements: This work was supported by NIH grant HG00022. We thank H. Lehrach for providing us with ICRFc102A07119 .

7. D11S35- Nucleic Acids Research (1990), Vol. 18, No. 19, pg. 5921

Dinucleotide repeat polymorphism at the D11S35 locus

M.Litt*, V.Sharma and J.A.Luty
Oregon Health Sciences University, Portland, OR 97201, USA

Source and Description: Phage 2−22 was from a human genomic library of CF52, a mouse X human somatic cell hybrid that retained a single translocated human chromosome, t(11;16)(q13;p11) on a rodent background (1). A Sau3A subclone of this phage that contained a $(GT)_{17}$ repeat was sequenced and sequences flanking the repeat (EMBL accession no. X52579) were used to design PCR primers.

PCR Primers:
(#780)-5′-ACAATTGGATTACTACTAGC-3′
(#781)-5′-TGTATTTGTATCGATTAACC-3′

Polymorphism: Allelic fragments were resolved on DNA sequencing gels. Lengths of allelic fragments (nt) were: A1=162, A2=160, A3=158, A4=156, A5=154, A6=152.

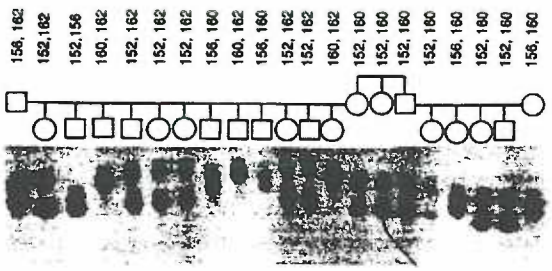
Frequencies: Allele frequencies in 17 unrelated European Caucasians: A1=.12, A2=.26, A3=.15, A4=.12, A5=.18, A6=.18. Heterozygosity=0.88; PIC=0.79.

Chromosomal Localization and Mendelian Inheritance: In situ hybridization and hybridization of phage 2−22 to genomic DNAs from a somatic cell hybrid panel indicated localization to 11q22 (1). Linkage analysis of the D11S35 MspI RFLP in CEPH families supports the order cen-D11S84-D11S35-D11S424-qter (2). Mendelian inheritance was observed in three informative families with a total of 23 children.

PCR Conditions: We carry out PCR in a total volume of 25 μl containing: 50 ng genomic DNA, 25 pmoles of each primer, 1.5 mM $MgCl_2$, 200 μM unlabeled dNTPs, 50 mM KCl, 10 mM Tris-Cl^-, pH 8.3, 0.6 units Taq polymerase (Perkin-Elmer/Cetus) and 0.01% gelatin. 1 μCi [^{32}P]dCTP is added to each sample. Amplification is for 30 cycles with denaturation at 93°C for 1 min, annealing at 40°C for 2 min and extension at 72°C for 2 min. Products are resolved on DNA sequencing gels and visualized by autoradiography.

Acknowledgements: This work was supported by Grant GM-32500 from the National Institutes of Health.

References: 1) Maslen,C.L., Jones,C., Glaser,T., Magenis,R.E., Sheehy,R., Kellogg,J. and Litt,M. (1988) *Genomics* 2, 66−75. 2) Julier,C., Nakamura,Y., Lathrop,M., O'Connell,P., Leppert,M., Litt,M., Mohandas,T., Lalouel,J.-M. and White,R. (1990) *Genomics* 7, 335−345.



* To whom correspondence should be addressed

C: Chromosome 21 linkage map

Seven microsatellite markers have been used to construct a linkage map of chromosome 21. Table 2 lists the PIC, heterozygosity, number of informative meioses, number of alleles and allele frequencies for the polymorphisms which were used to construct the map. An average of 65 CEPH parents were typed for each locus (range 46-74). The PIC and heterozygosity averaged 0.68; the heterozygosity ranged from 0.32 to 0.90 and the PIC from 0.56-0.83. The average number of informative meioses for each locus was 370 (range 185-540). The alleles for three loci were recoded into 4 allele systems. This reduces the apparent PIC of the loci but not the heterozygosities, which the computer program CRIMAP calculates directly from the data. Table 3 shows the two point linkage data between markers. This data was used by the "build" option of CRIMAP to order loci. The map length for D21S172-D21S198 is 56.5 cM when averaged between the sexes (Table 4). The male map has a length of 42.1 cM and the female map of 72.9 cM (Table 4). All loci except D21S236 are ordered with odds exceeding a million to one against permutation of adjacent loci as calculated by the Flips3 option of CRIMAP (Table 5). D21S236 could not be ordered in relation to D21S172 with odds exceeding 1000:1, though a proximal position for D21S236 is favored (LOD score = 2.3). On the sex averaged map the average distance between loci is 11.3 cM. The largest distances are between D21S11-D21S218 (17cm) and D21S218-D21S167 (16cM). The most striking difference in recombination frequencies between the sexes is in the interval bounded by D21S172 and D21S11, where the female map is >5 fold larger (16.4 cM) than the male map (3.1 cM).

TABLE 2: Data for Microsatellite Systems:

System	#Ind	Hetero*	PIC	IM+	Allele/Freq		
D21S172|pcr	70	0.71	0.70	427	1/0.43	2/0.33	3/0.19
					4/0.05		
D21S167|pcr	53	0.64	0.80	269	1/0.31	2/0.41	3/0.25
					4/0.03		
D21S168|pcr	74	0.77	0.74	486	1/0.03	2/0.05	3/0.20
					4/0.39	5/0.18	6/0.07
					7/0.03	8/0.06	
D21S236|pcr	66	0.65	0.67	363	1/0.08	2/0.08	3/0.26
					4/0.11	5/0.02	6/0.01
					7/0.45		
D21S218|pcr	73	0.42	0.56	183	1/0.23	2/0.06	3/0.56
					4/0.15		
D21S198|pcr	46	0.80	0.81	320	1/0.26	2/0.27	3/0.25
					4/0.13	5/0.05	7/0.03
D21S11|pcr	72	0.90	0.83	540	1/0.01	2/0.01	3/0.13
					4/0.03	5/0.07	6/0.11
					7/0.03	8/0.22	9/0.22
					10/0.15	11/0.01	12/0.01

* Heterozygosity is calculated directly from percent heterozygotes:

+ Informative meioses

TABLE 3: Two-Point LOD Scores

Numbers in the table below correspond to these probe-enzyme systems:

0	D21S172	1	D21S167	2	D21S168
3	D21S236	4	D21S218	5	D21S198
6	D21S11				

Bottom diagonal is recombination fraction; top is LOD score

	0	1	2	3	4	5	6
0		2.14	4.38	43.80	7.95	4.13	47.41
1	0.35		25.77	0.68	8.82	20.61	4.52
2	0.33	0.08		3.17	11.63	45.86	7.50
3	0.05	0.41	0.34		4.04	0.54	40.72
4	0.22	0.14	0.16	0.27		3.29	15.96
5	0.32	0.10	0.06	0.43	0.26		2.99
6	0.09	0.31	0.32	0.09	0.14	0.37	

TABLE 4: Chromosome 21 Linkage Maps

Option chosen: build

	LOCUS	SUBCLONE
0	D21S172	1H9
1	D21S167	
2	D21S168	
3	D21S236	JA5
4	D21S218	JB2
5	D21S198	VS#1
6	D21S11	VS17TB2

4A: Sex averaged map (recomb. frac., Kosambi cM):

0	D21S172			0.0
		0.09	9.2	
6	D21S11			9.2
		0.17	17.3	
4	D21S218			26.5
		0.16	16.5	
1	D21S167			43.0
		0.08	7.7	
2	D21S168			50.7
		0.06	5.8	
5	D21S198			56.5

4B: Sex-specific map (recomb. frac., Kosambi cM -- female, male):

0	D21S172			0.0			0.0
		0.16	16.4		0.03	3.1	
6	D21S11			16.4			3.1
		0.26	29.2		0.10	10.2	
4	D21S218			45.6			13.3
		0.14	14.5		0.15	15.7	
1	D21S167			60.1			29.0
		0.05	5.0		0.09	9.5	
2	D21S168			65.1			38.5
		0.08	7.8		0.04	3.6	
5	D21S198			72.9			42.1

TABLE 5: Likelihoods for Permutation of Adjacent Loci

Option chosen: flips3

3 D21S236
0 D21S172
6 D21S11
4 D21S218
1 D21S167
2 D21S168
5 D21S198

number of loci to flip = 3

Original order, & its log10_likelihood, followed by flipped orders, with their relative log10_likelihoods (= log10_like[orig] - log10_like[curr])

3	0	6	4	1	2	5	-300.80
6	-	3	-	-	-	-	10.85
6	3	0	-	-	-	-	13.68
0	6	3	-	-	-	-	24.99
-	6	0	-	-	-	-	26.31
0	3	-	-	-	-	-	2.29
-	4	-	0	-	-	-	49.22
-	4	0	6	-	-	-	54.00
-	6	4	0	-	-	-	48.37
-	-	4	6	-	-	-	29.18
-	-	1	-	6	-	-	86.20
-	-	1	6	4	-	-	80.53
-	-	4	1	6	-	-	86.74
-	-	-	1	4	-	-	17.85
-	-	-	2	-	4	-	31.86
-	-	-	2	4	1	-	34.66
-	-	-	1	2	4	-	29.39
-	-	-	-	2	1	-	10.72
-	-	-	-	5	-	1	9.33
-	-	-	-	5	1	2	16.54
-	-	-	-	2	5	1	4.00
-	-	-	-	-	5	2	6.13

IV: DISCUSSION

A: Chromosome 21

A number of important markers have been developed from chromosome 21 and a number of objectives realized.

1. Using the microsatellites from chromosome 21, we have constructed a linkage map which covers 75-85% of the genetic distance of this chromosome. Our map is considerably shorter than previous maps of chromosome 21 and compares favorably to a preliminary index map constructed for the US human genome project by Dr. Antonarakis. (Antonorakis *et al.*, 1992; Appendix C)

2. We were invited by Dr. Antonorakis to contribute our data for incorporation into two maps which are being constructed for the US genome initiative. D21S11 & D21S167 are included in a preliminary index map of chromosome 21 which consists of 13 markers. This index map is part of a collaborative effort by many laboratories to construct a complete map of the human genome. Coordinated and assembled by Dr. Donis-Keller, a map of the human genome map has been submitted to *Science* for publication. A comprehensive map of chromosome 21 which consists of 38 polymorphic markers genotyped in CEPH families by PCR is also being constructed by Dr. Antonorakis for publication at a later date. (MG McInnis *et al.*, 1992; Appendix B)

3. Four microsatellites from the proximal long arm of chromosome 21 were successfully developed from areas traditionally poor in genetic markers.

4. I used the microsatellites from D21S11 and D21S13 to determine the genotypes of 166 individuals from FAD families, which were incorporated into a linkage study of this disease by Dr. Schellenberg (Kamino *et al.*, 1992; Appendix E).

Comparison of chromosome 21 linkage maps

Petersen *et al.*'s (1991) chromosome 21 linkage map is currently the most comprehensive of this chromosome. Various features have been summarized in the

introduction: *i.e.* 25/27 markers are RFLPs; only three markers (D21S13, D21S110 & D21S11) are from the proximal long arm of the chromosome; the observed heterozygosity of the loci used in the map averages 0.45 and ranges from 0.11 to 0.76. This map covers most of the long arm of chromosome 21. D21S13 is the most proximal marker and COL6A1 is the most distal marker. When no interference is assumed, the map length using 27 markers is 143 cM in males, 182cM in females and158 cM when sex averaged. Petersen *et al.* also constructed an anchor map consisting of eight highly polymorphic loci spaced 10-20 cM apart. The anchor map was 87.4 cM in males, 122.6 cM in females and 102.3 cM when sex averaged. The anchor map and our microsatellite map were made using the Kosambi mapping function.

One locus is shared between our map and Petersen's map- D21S11. Using the two point linkage data as reported in our manuscripts, we note that D21S172 is within 1 cM of D21S13 and that D21S198 is linked to PFKL at theta= O. From this we deduce that our map covers approximately 75-85% of Petersen's map which has a length nearly 1.5 times longer than ours. The portion of chromosome 21 which is not covered by our map is the telomeric region distal to CRYA1/PFKL. On Petersen's sex averaged anchor map this represents about 25 cM. D21S198 is in the "middle" of q22.3, which from previous studies contains 50% of the genetic map of chromosome 21. This is consistent with our observation that our map covers approximately 25% less than Petersen's which includes the most distal loci.

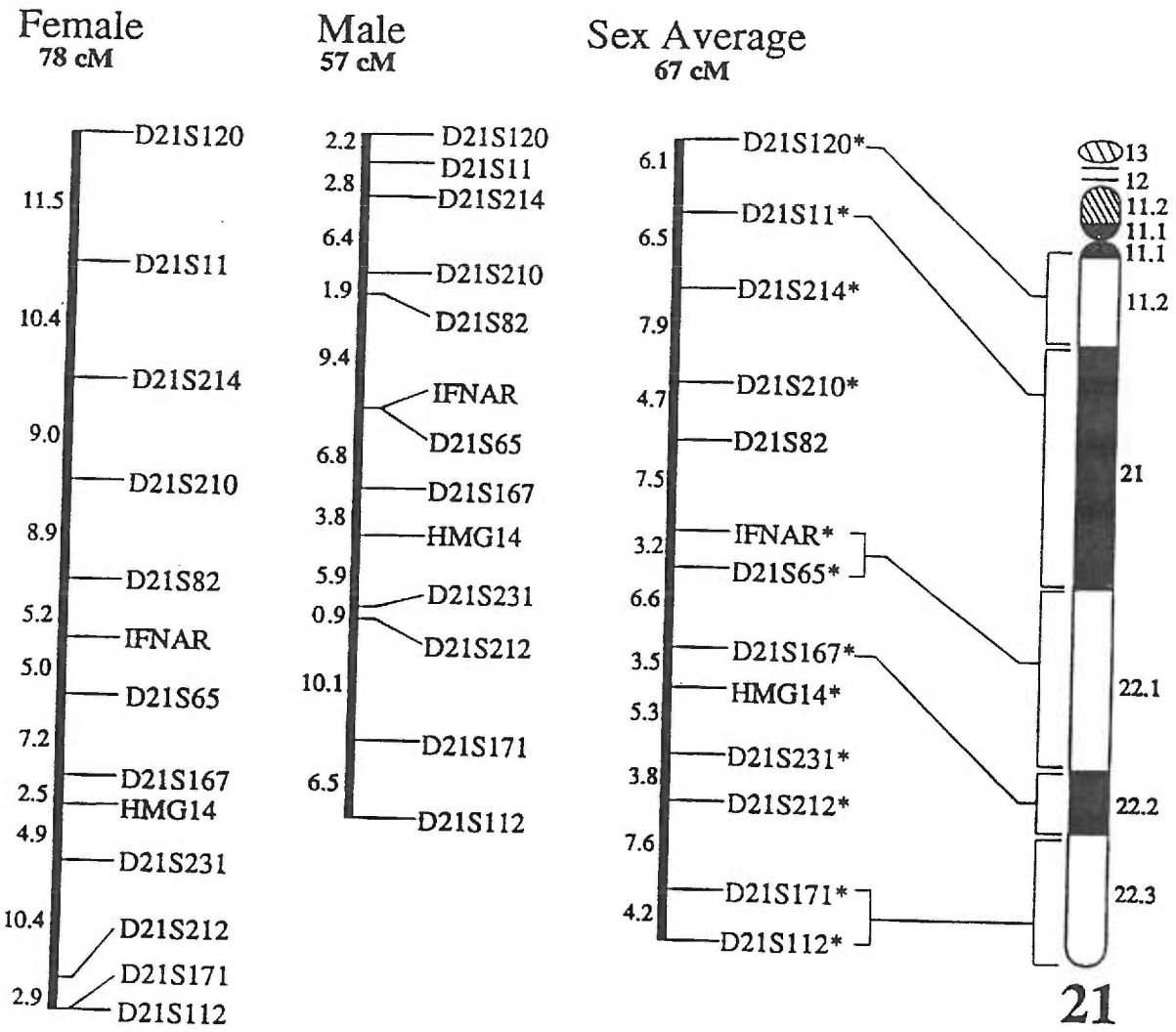
Both maps show that recombination in female meioses is significantly more frequent than in males. On our microsatellite map the ratio of the map length between females and males is 1.73 (72.9 cm÷ 42.1cM); on Petersen's anchor map the ratio for the corresponding interval is 1.60 (103cM÷64cM). This ratio is not constant along the length of the chromosome but reflects areas of variability (Petersen *et al.*, 1991; Cox and Shimizu, 1991). The most striking difference is in 21q proximal to D21S11. In our map the recombination frequency between D21S172 (essentially D21S13) and D21S11 is 0.03 in males, 0.16 in

females and 0.09 when sex averaged; the female to male ratio is 5.30. On Petersen's anchor map the recombination frequency between D21S13 and D21S11 is 0.05 in males, 0.20 in females and 0.14 when sex averaged; the female to male ratio is 4.0. Both maps show an interval where male recombination is twice that of females. In our map this interval is between D21S168 and D21S167; in Petersen's map, it is bounded by MX1 and CRYA1.

The data from our map is being incorporated into two maps being constructed for the US genome project by Dr. Antonorakis at Johns Hopkins University School of Medicine. These maps have not yet been published. An index map consisting of 13 markers with heterozygosities which range from 60% to 90% has a length of 57 cM in males, 78 cM in females and 67 cM when sex averaged (Figure 6). Twelve markers have heterozygosities >0.70; the average spacing between markers on the sex averaged map is 5.6 cM; the D21S11 tetramer and D21S167 were incorporated into this map. Dr. Antonorakis estimates that this map covers 95% of the physical length of chromosome 21 and predicts that "the success of mapping any disease locus that may reside on chromosome 21 is a virtual certainty, even with relatively modest pedigree resources" (Antonorakis *et al.*, 1992; Appendix C). A comprehensive map of chromosome 21 is also being constructed. Preliminary data sent to us shows that this map consists of 38 loci including 9 genes; these are primarily (CA)n, (AAAT)n and (GATA)n microsatellites. 24 markers have heterozygosities >0.64; 26 loci were ordered with odds exceeding 1000:1. The total map length (D21S215-D21S171) is 77.9 cM sex averaged, 59.5 cM in males and 97.0 cM in females. The average distance between loci is 2-3 cM (McInnis *et al.*, 1992; See appended abstract). We have not received further information which shows the position of our microsatellites relative to the new markers which define the proximal and distal ends of this map.

It is notable that the genetic linkage maps constructed with microsatellites are considerably smaller than previously published linkage maps. This is probably due to a significant number of errors (false recombination events) being incorporated into the earlier maps which results in inflated genetic distances. (Antonorakis *et al.*, 1992) As more loci are

Figure 6: The index map for chromosome 21. The position and distances between markers are shown relative to an ideogram of chromosome 21. * denotes markers which can be amplified by PCR. (From Antonorakis *et al.* 1992; see Appendix C)


Female
78 cM
D21S120
11.5
D21S11
10.4
D21S214
9.0
D21S210
8.9
D21S82
5.2
IFNAR
5.0
D21S65
7.2
D21S167
2.5
HMG14
4.9
D21S231
10.4
D21S212
D21S171
2.9
D21S112
Male
57 cM
2.2
D21S120
D21S11
2.8
D21S214
6.4
D21S210
1.9
D21S82
9.4
IFNAR
D21S65
6.8
D21S167
3.8
HMG14
5.9
D21S231
0.9
D21S212
10.1
D21S171
6.5
D21S112
Sex Average
67 cM
6.1
D21S120*
6.5
D21S11*
D21S214*
7.9
4.7
D21S210*
D21S82
7.5
3.2
IFNAR*
D21S65*
6.6
3.5
D21S167*
HMG14*
5.3
D21S231*
3.8
D21S212*
7.6
4.2
D21S171*
D21S112*
13
12
11.2
11.1
11.1
11.2
21
22.1
22.2
22.3
21

typed the chance for making mistakes increases. Our map has not relied on genotypes determined in other laboratories and has been checked for errors in data entry or film reading. Using the chrompic option of CRIMAP, the grandparental origins of alleles and the number and locations of recombinations for each child's chromosome were identified. The genotypes for chromosomes displaying two recombination events between closely spaced (adjacent) loci were re-checked by examining the original autoradiographs. No changes were made to the data unless an obvious error was detected.

Other estimates of the genetic length of chromosome 21 have been calculated from meiotic chiasma counts in males (Petersen *et al.*, 1991; Laurie & Hulten, 1985). The mean chiasma frequency was 1.05 which corresponds to a map length of 53 cM for males. This is considerably smaller than Petersen's estimate but very close to that determined by the maps constructed with microsatellites.

Contributions to the index map of chromosome 21:

In total, I have been developed 8 repeats from chromosome 21: 6 have heterozygosities >0.7 and are suitable for use as index markers; 7 markers have been typed through 40 CEPH families; 2600 informative meioses were examined. The current CEPH V5.0 data base for chromosome 21 does not contain any microsatellite polymorphisms so these are an important addition. Also, the CEPH genotypes for the markers have been incorporated into two maps being constructed for the US human genome initiative. HGM 11 lists 10 markers with PIC's >0.5 which have been mapped to specific regions of chromosome 21; 5 were developed in this study.

Four new microsatellites were successfully developed from the proximal half of 21q (q11.2-21), a region generally poor in genetic markers and which remains implicated in FAD. Two microsatellites were isolated from clones identified with single copy probes; the remaining were subcloned from a cosmid mapped by in situ hybridization to q11.2, and a previously unmapped cosmid. Three have heterozygosities >0.7. These are:

1) The (AGAT)n tetramer microsatellite from D21S11 has significantly improved the informativeness of this locus. Dr. Antonorakis, the chairman for the US genome chromosome 21 index map, has stated that this is the only index marker in the region which can be characterized by PCR (Antonorakis, pers. comm.). The primer sequences for PCR amplification were provided to T. Siddique who has indicated that this locus may be the closest to the familial form of amyotrophic lateral sclerosis (Siddique, pers. comm.). D21S11 is a reference locus on chromosome 21 and is linked to FAD.

2) D21S236 and D21S172 are highly informative markers from the most proximal end of 21q (q11.2). D21S236 appears to be centromeric to D21S172, but could not be ordered with odds greater than 1000:1. When the data from CEPH for the Taq1 RFLP at D21S13 was incorporated into our analysis, it could not be ordered in relation to D21S172 and D21S236. The most likely position for D21S13 is between D21S236 and D21S172 but the odds in favor were only about 60:1. Haines *et al.* tentatively placed D21S172 proximal to D21S13 on a map which he constructed. (Haines, pers. comm.) While we cannot confirm this, it appears safe to conclude that D21S236 is more proximal than D21S13/S16 which was the most proximal marker on Petersen's *et al.* linkage map. The recombination distance between D21S236 and D21S13 is approximately 2 cM.

3) Although a microsatellite has been cloned from D21S13, a targeted locus, it is poorly informative. A haplotype consisting of this microsatellite and the Taq1 RFLP at D21S13 can be employed to increase the PIC of the locus to 0.69. Another D21S13 microsatellite we cloned was left uncharacterized once Van Broekhoven published a moderately informative CA repeat from the same cosmid (Stinsson & Van Broeckhoven, 1991). D21S172 is within 1 cM of D21S13 and may serve as a substitute locus for linkage studies.

FAD Study: I used the D21S11 tetramer and two microsatellites from D21S13 (Stinsson & Van Broeckhoven, 1991; Guo *et al.*, 1990) to type 166 individuals from FAD families

collected at the Alzheimer's Disease Center of Oregon by Dr. H. Payami and at the University of Washington by Dr. G. Schellenberg. These individuals were typed by PCR, in part because there was insufficient genomic DNA to make Southern blots (Schellenberg pers. comm.) and because the microsatellites are more informative than the associated RFLPs. Dr. Schellenberg incorporated these genotypes into a large study which examined the frequency of APP mutations in individuals with FAD and linkage of FAD to chromosome 21 markers. The linkage analysis confirmed earlier results of Schellenberg's by failing to find linkage between proximal 21q markers (including APP) and FAD. In addition, Shellenberg found that APP mutations postulated to be causative of FAD (Goate *et al.*, 1991), were not present in unrelated affected individuals from 130 families which he studied (Kamino *et al.*, 1992) The manuscript is appended.

B: Characteristics of Microsatellites

The Abundance of microsatellites

Simple sequence repeats are an abundant source of polymorphisms used to construct linkage maps. To determine whether exclusively characterizing (CA)n dinucleotide repeats reduces our efficiency in isolating useful polymorphisms, we compared the frequency of CA repeats to those of selected tri and tetramer repeats. We did this by screening a flow sorted chromosome 21 cosmid library with an oligonucleotide pool consisting of equimolar amounts of 7 different repeats (Table 7A) and with the CA/GT copolymer (Figure 7). Since the base composition of the oligomer probes vary depending on the repeat motif, the oligomers in the pool will have different melting temperatures. Three molar tetramethyl ammonium chloride (TMACl) was used to create base composition independent hybridization conditions where probe length is the sole determinant of the Tm of the hybrids (Wood *et al.*, 1986). To determine what proportion of cosmids positive for CA repeats contained multiple CA repeats, Southern blots were prepared from cosmids digested with AluI, Sau3A, & HaeIII (Figure 8). These enzymes have 4 base recognition sequences which

cut more frequently and generate shorter fragments than restriction enzymes based on longer recognition sequences, and are thus more likely to separate CA repeats on a single cosmid. The efficiency of the labeling reaction for the oligo pool was 40%. Southern blots, CA/GT copolymer and oligomer hybridizations are as detailed in Material and Methods.

Results: Of 576 cosmid clones grown, lifted and screened from replicas of six 96 well microtiter plates, 25% were positive with the CA repeat probe. Twenty five CA positive cosmids were digested with AluI, Sau3A, & HaeIII: 11 (44%) had one repeat, 9 (36%) had two repeats and 5 (20%) had three repeats. The 96x96 array was more difficult to score but a conservative estimate based on counting randomly choosen rows of 96 or 10x10 squares showed that approximately 20-25% of cosmids contained a CA repeat. (Fig. 7) Using Table 7C we calculate that 100 CA positive cosmids will represent 176 CA repeats. Taking 45 Kb of genomic DNA as the average insert size of a cosmid, there is one CA repeat every 25 Kb. However, since only 25% of cosmids are CA positive this equates to one repeat every 100 Kb. By extrapolation, we conservatively estimate that there are 33,000 CA repeats in the human genome. (3,300,000 Kb/haploid human genome ÷ 100Kb/repeat=33,000 repeats.)

Sixteen percent of the clones from the microtiter plates are positive with the oligomer pool. A conservative estimate for the 96x96 array shows that 6-10% of the cosmids hybridized to the pool. By using the larger firgure of16%, the frequency of individual non CA STR's in the pool is approximately 2.3% or 1 repeat per 2 megabases (Mb) of DNA. The actual representation in the library of each individual repeat is unknown. The tetramer AATC has been reported to be particularly abundant and well distributed (Edwards *et al.*, 1991). We screened the 96x96 cosmid array with an $(AATC)_7$ oligomer probe in order to independently determine it's frequency which was found to be 0.7-1.5 % or one repeat every 3-6 Mb of DNA.

Table 7A: Composition of oligonucleotide pool

(TAAA)7	(TTA)9
(TTCA)7	(TAG)9
(ACAG)7	(AGC)9
(AGAT)7	

Table 7B: % of CA & non CA STR positive clones from Lehrach's flowsorted chromosome 21 library.

Repeat	96x96 array*	Colony lifts(=576clones)
% CA+ clones:	20-25%	25%
% non CA STR+ clones:	6-10%	16%
% AATC+ clones:	0.7-1.5%	nd

* Gridded array of 9216 cosmids clones on a single 22x22 cm membrane

Table 7C: # of CA repeats in 25 CA positive cosmids:

# of repeats	1 CA+	2 CA+	3 CA+
# of cosmids (total=25)	11 (44%)	9 (36%)	5 (20%)

Figure 7: Lehrach's 9216 clone array of a chromosome 21 flow-sorted cosmid library contained on a 22 cm X 22 cm nylon membrane and screened with CA/GT copolymer.

Lehrach 96 x 96 array probed with CA/GT copolymer

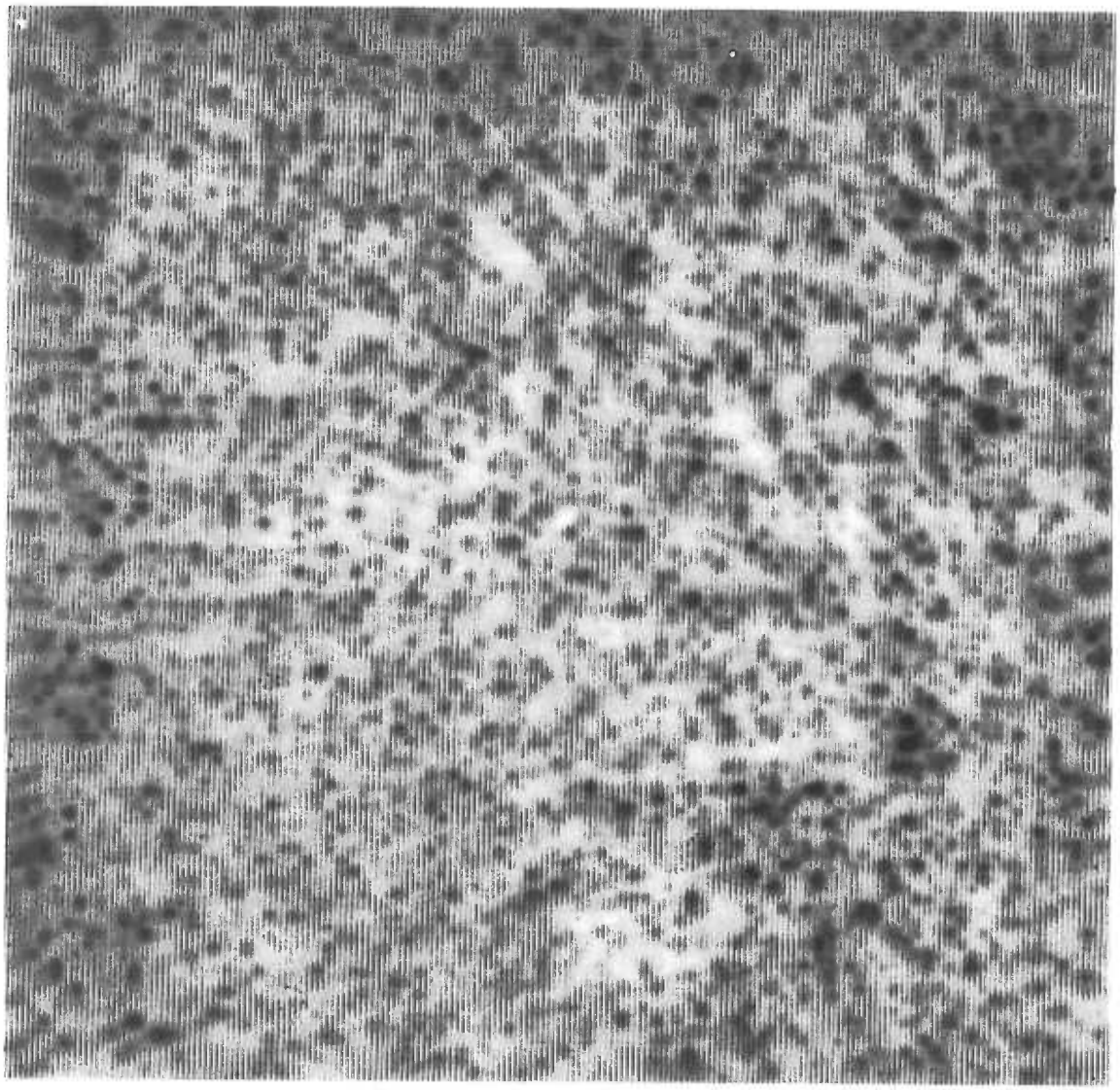
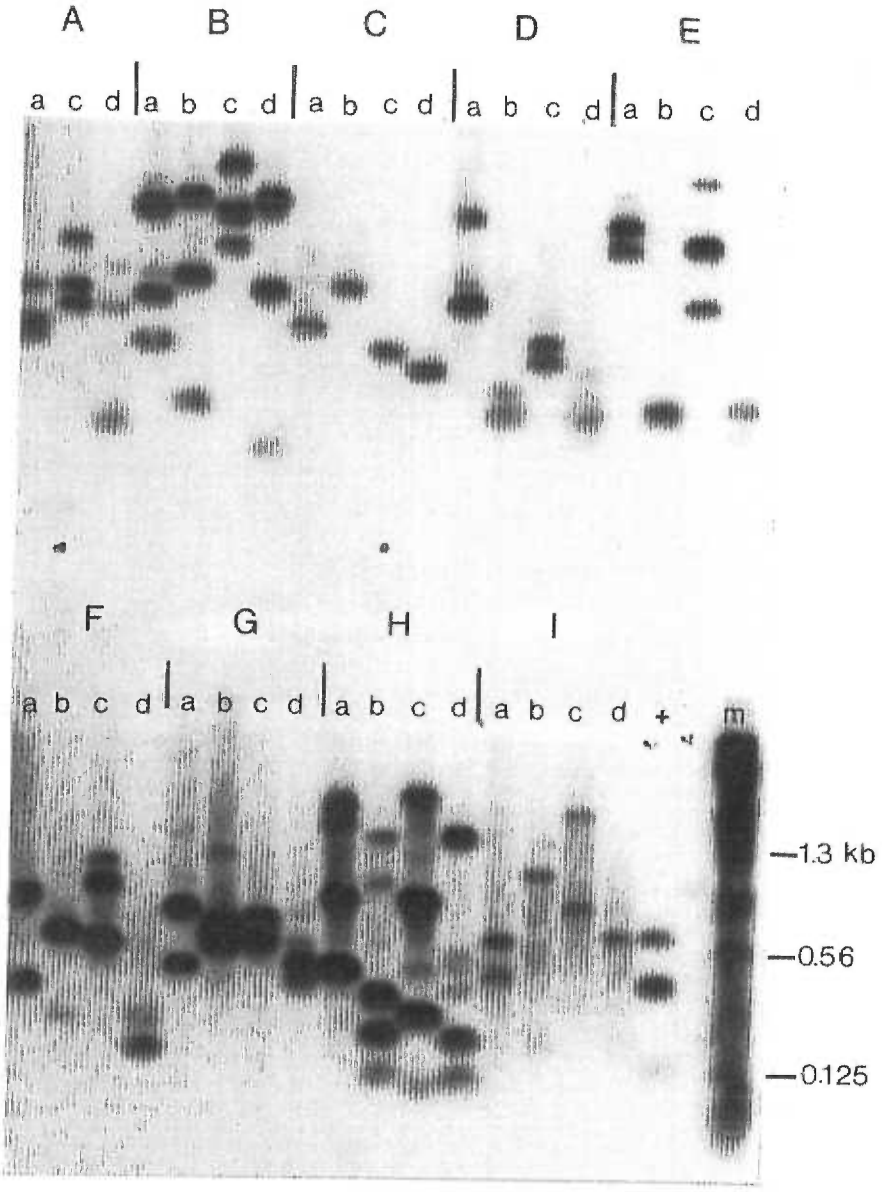


Figure 8: Southern blot of 9 CA repeat positive cosmids (A-I) from Lehrach's flow sorted chromosome 21 library digested with Sau3A (a), Alu1 (b), HaeIII (c) and Alu1 & HaeIII (d); and probed with CA/GTcopolymer. A size standard (m) is shown for the lower panel. Cosmid B (D21S116) contains 3 CA repeats; comsid C (D21S172) contains one CA repeat.


A
B
C
D
E
a c d a b c d a b c d a b c d a b c d
F
G
H
I
a b c d a b c d a b c d a b c d + m
—1.3 kb
—0.56
—0.125

Discussion: Since the discovery of CA microsatellites, other short repeat motifs of 3-6 nucleotides have been examined and found to be polymorphic. A recent debate in the genetics community that uses and developes microsatellites caused us to examine our repeat screening strategy which focuses on isolating CA repeats exclusively. Namely, some groups feel that the close spacing of alleles and "shadow" bands typical of CA repeats interferes with the accurate determination of genotypes. It has been suggested that tri and tetramer repeats are easier to use, and that certain motifs such as the tetramer AATC, are nearly as abundant and polymorphic as CA microsatellites (Edwards *et al.*, 1991).

Our results confirm the abundance of CA repeats relative to individual tetramer and trimer repeats which are much less abundant. The number of CA repeats in our estimate is similar to those cited in the literature (Weber, 1990). Our estimate for the frequency of other non CA STRs , in particular AATC is much less than what has been previously reported (Edwards *et al.*, 1991). Edwards screened 1020 lambda phage (average insert size 15 Kb) from an X chromosome genomic library with individual 30 bp oligonucleotides for 5 different repeats: 3-9% were positive for each repeat allowing him to estimate that they occurred every 170-500 Kb. The highest frequency we calculate is at least 4 fold less or 1 non CA STR every 2 Mb.

The degree to which this difference in reported non CA STR frequencies reflects differing hybridization conditions is unclear. Do the hybridization conditions vary in discriminating between perfect and interrupted repeats, long and short repeats or numbers of repeats within clones? Are we missing them in our screen? Our hybridization signal displays a full range of intensity which does not reflect differences in the amount of DNA on the array since the background is known to be fairly uniform on the basis of screening the array with labeled vector. This suggest that we are seeing differences in clones due to repeat copy number and perhaps length or type of repeat. A number of clones would need to be sequenced in order to confirm this.

While STRs may be less abundant than reported, this may not seriously erode the potential for finding them when using oligo pools. However, unlike CA repeats, there is less information available concerning the proportion of non CA STRs which are polymorphic. Edwards states that the minimum length of non CA STR to obtain a hybridization signal under his conditions is 7 repeat units and suggests that this is long enough to insure that 50% of the clones characterized will be polymorphic. Of 17 STRs characterized by Edwards, 8 had one allele; they included long and short, tri and tetrmer repeats. Of the remaining nine which were polymorphic, the heterozygosities ranged from 19-91% (55% average). By comparison, 45% of the CA repeats developed in our laboratory have PIC's over 0.7. The median PIC for 51 repeats (including 9 monomorphic loci) was 0.65. If we apply this percent to the total number of CA repeats estimated to be in the genome, there are potentially 15,000 highly informative repeats. Our estimate of 33,000 CA repeats with perhaps 50% being polymorphic is remarkably close to that of Weber (Weber, 1990).

There are practical issues to consider in the choice of repeat to use for screening, such as whether one is developing polymorphisms from a small number of clones or out of a genomic library. If one has a limited number of clones from a target region for which a good CA repeat polymorphism could not be developed, it can be worthwhile to look for non CA STRs. This approach led to the development of the D21S11 tetramer repeat. As with CA repeats, it is likely that the characteristics of a particular STR (such as "ease of use", PIC) will be locus specific. Dracopoli, who recently developed 100 CA repeats on chromosome 1, found 8 non CA STRs which he did not charcterize due to "poor amplification, extreme stuttering or the amplification of multiple secondary bands" (Hudson *et al.*, 1992). Finally, it is important to note that many of the typical frustrations researchers have experienced with scoring CA microsatellite alleles may be avoided, in part as a direct result of the techniques demonstrated in our laboratory. Our experience suggests

that there is no need to switch from primarily developing CA repeat polymorphisms to developing other non CA STRs.

An additional benefit to screening libraries with CA repeat probes is that it serves to give an indication of the quality of the library. For example, the first library we screened was sent to us by P. de Jong and consisted of flow sorted chromosome 21 DNA cloned into lambda phage. Upon screening approximately 10,000 clones, less than 0.1% were CA repeat positive which is very far removed from our expectation based on literature citations and our previous experience with screening chromosme 11 and X specific phage libraries. When this information was relayed to de Jong, he subsequently confirmed that the library was unrepresentative of human sequences as a result of being heavily contaminated with *Pseudomonas* DNA (Yokobata, pers. comm.). In another case, a microdissected chromosome 20 plasmid library constructed by Lee Allen was screened with a CA probe by Marcus Grompe as part of his characterization. By comparing the frequency of repeats in the library to that expected based on the known abundance of CA repeats, he concluded that the library was representative of human sequences. (Grompe. pers. comm.)

In summary, CA repeats are found in 25% of cosmids in Lehrach's flow sorted library (of which 50% contain two or more repeats) and are spaced roughly every 100 kb of genomic DNA. Non CA STRs were found in 2.3 % of cosmids or 1 repeat every 2 Mb. There are 44 unique tri and tetramer repeat motifs (Edwards *et al.*, 1991) If this data is extrapolated to the entire human genome, there are consevatively 33,000 CA repeats and 69,000 tri & tetramer repeats which would mean a simple sequence repeat every 31 Kb of human genomic DNA . This figure is somewhat different from that calculated by Tautz and Edwards who estimated that a repeat is present every 10-20 Kb in the human genome (Tautz 1989; Edwards *et al.*, 1991). Though non CA STRs constitute a substantial source of potential polymorphisms, we conclude that CA repeats alone should prove sufficiently abundant to generate the polymorphisms necessary for most linkage map constructions.

Length of CA microsatellites vs heterozygosity

Table 8 summerizes the heterozygosity, sequence and length of the microsatellite repeats as well as the probe which was used to indentify the subcloned allele. The probes were either the CA/GT copolymer or the $(CA)_{15}$ oligomer.

Table 8: Length and sequence of repeats

Locus	Het.	#Alleles	Probe*	Repeat sequence of subcloned allele
D21S11	0.90	>14	o	(TCTA)4(TCTG)6(TCTA)3(ATCT)10
D21S198	0.81	9	c	(TG)23
D21S167	0.80	13	c	(GT)19
D11S35	0.79	6	c	(GT)17
D21S168	0.77	8	c	(GT)19
D1S117	0.77	11	c	(TG)21(TA)10
D9S55	0.75	9	c	(TG)24
D14S43	0.72	10	c	(TG)14-N29-(TG)10
D21S172	0.71rev	8	c	(CA)20
D21S236	0.71	7	o	(TG)11
D4S233	0.69	6	o	(TG)13
D18S37	0.65	6	c	(AC)15
D1S116	0.62	7	c	(GT)17
D21S218	0.55	4	o	(TG)13
D21S13	0.42	5	c	(AC)18
#1B	0.33	2	o	(GT)11
1H8	0.30	3	c	(TG)14-GC(GT)5
2D9	0.10	2	c	(GT)>15
D11S286	0.00	1	c	(TG)21

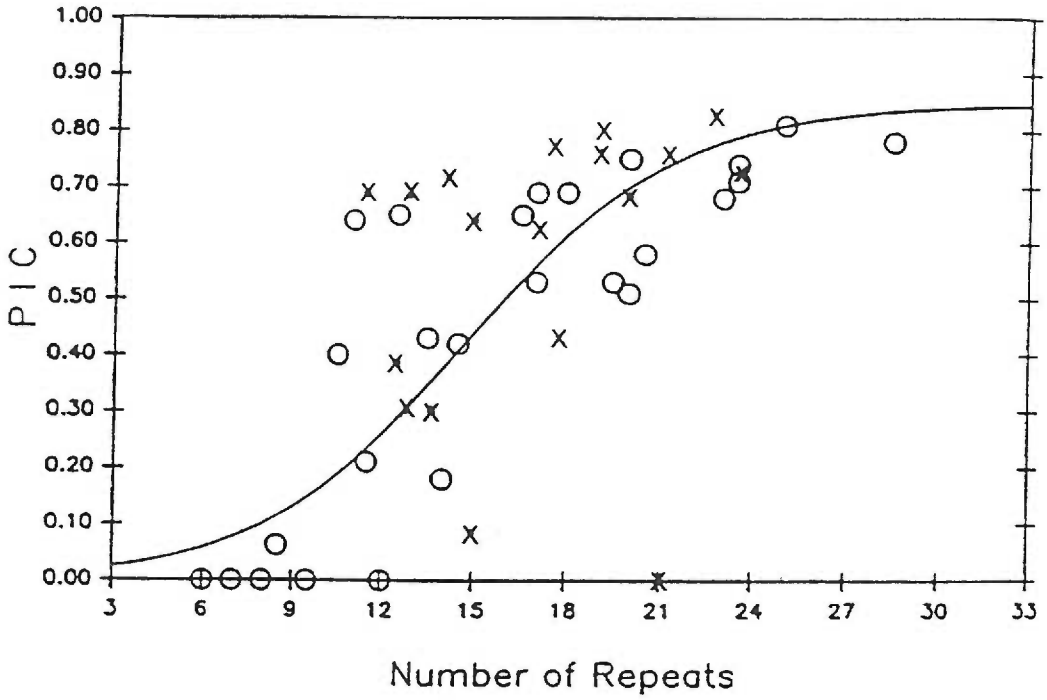
* o= $(CA)_{15}$ oligo probe; c= CA/GT copolymer

J. Weber has postulated that the PIC of CA microsatellites is related to their length (the average number of repeat units at a locus), with longer CA repeats (>15 units) being more polymorphic than short ones (Weber, 1990). This is important for practical reasons concerning the choice of screening methods used to identify repeats since it is preferable to

bias the selection towards clones most likely to be polymorphic. While inspection of Table 8 and Figure 9 appears to generally confirm this principle, our data and that of others (Hudson *et al.*, 1992) has shown that this correlation is not absolute. Unlike Weber, we have found long repeats which are monomorphic (D11S286). Repeats between $(CA)_{15}$ and $(CA)_{11}$ in length are variable in PIC with some short repeats being quite polymorphic (D21S236). What does appear to be most true is that the PIC is very much reduced for repeats less than 10 units long. Our original repeat screening protocol used long (>100 nt.) CA/GT copolymer sequences, which are not expected to hybridize stably to repeats less than 12 repeat units long (Weber, 1990). Thirteen CA repeats were identified with the copolymer probe and averaged $(CA)_{18}$ in length, with the shortest simple (uninterrupted) repeat being D18S37 (CA_{15}). Eight of the repeats identified with the copolymer are index quality. For reasons of convenience, the $(CA)_{15}$ oligo probe was used to screen subclones since the probe labeling procedure is simple, no prehybridizations of filters are necessary, and the hybridization and wash times are short. Six repeats were sequenced (including a $(GT)_9$ and an interrupted $(CA)_{13}$ which are not listed in the table); the average length was $(CA)_{12}$. Repeats as short as $(CA)_9$ but none longer than $(CA)_{15}$ were identified. Only one microsatellite identified with $(CA)_{15}$ oligo probes had a heterozygosity over 0.70. Other researchers in our lab have also identified numerous short or interrupted repeats with low PICs when using $(CA)_{15}$ oligomers as a probe (Browne, Hauge pers. comm.). It is unlikely that this is coincidence and appears to be inherent to the use of oligo probes. Based on observations by Ostrander and Browne, the stringency of hybridizations can be increased to a point which discriminates between perfect versus interupted repeats. Since interrupted repeats tend to be less polymorphic than uninterrupted repeats (Weber, 1990), this helps identify clones which should be preferentially charcterized. There is no data to suggest that the stringency of hybridization with $(CA)_{15}$ oligomers can be used to discriminate between repeats which are $(CA)_{15}$ (or less) in length and those that are considerably longer. For this purpose CA/GT copolymer appears to offer a reliable and

Figure 9: The informativeness of perfect CA repeat sequence polymorphisms as a function of repeat length. The plot was obtained by using the number of repeats found within the original sequences used for PCR primer sunthesis. Open circles are from Weber,1990; "X" are cosmids from Table 8.

PIC vs Repeat length



perhaps preferred method. In conclusion, there is an approximate correlation between the length of a repeat and it's potential for being polymorphic. Though useful polymorphisms were isolated using $(CA)_{15}$ oligomer probes, the clones identified by the CA/GT copolymer resulted in 8 of 10 of the high PIC index quality markers characterized in this dissertation.

New Alleles

In this study, genotypic incompatibilities within CEPH families were noted for 3 of 7 chromosome 21 loci typed through the 40 CEPH families (D21S11, D21S236, D21S168). These "allelic exclusions" can result from mispaternity, incorrectly scored autoradiogaphs/genotypes, sample mix-ups, or new alleles arising from mutations at the locus. The mutation rate of microsatellites in humans is not well defined. Some minisatellites have been shown to produce new alleles at rates of up to 5% per generation (Jeffreys *et al.*, 1988). These generally had heterozygosities which approached 1. With large well characterized data sets where mispaternity or mistyping can be excluded, the frequency of new microsatellite alleles may be calculable. Edwards did not note any genetic incompatiblities in 850 informative meioses with 5 non CA STRs (Edwards *et al.*, 1992). Kwiatkowski *et al.* identified two spontaneous mutations in the Venezualan reference families when they were constructing a linkage map of 9q with 15 microsatellite polymorphisms. The mutation rate they determined was .00045 per locus per gamete (Kwiatkowski *et al.* 1992). Litt and Luty have confirmed the presence of a new microsatellite alleles in CEPH families for various microsatellites. Dr. Litt estimates that they may be as frequenct as .001 per locus per genome. (Litt, pers. comm.) With my data, the frequency of detected allelic exclusions for the 9 markers typed through 40 CEPH families was 3 events in different individuals out of approximately 3200 informative meioses or .0009 new mutations per locus per gamete. The exclusions are in the process of being confirmed by retyping individuals with new aliquotes of primary genomic DNA stocks. They are in different families and have been duplicated in independent polymerase

chain reactions. As with other properties of microsatellites, the mutation rate may be locus specific and vary according to the PIC, repeat motif and length. The new mutations in this study were found at loci which have heterozygosities of 0.90, 0.77 and 0.71. With STR's new alleles may be indistinguishable from an allele already present in a family if they are the same size. In linkage analysis these individuals would appear as recombinants. The consequence of the error is to expand the apparent genetic distance between loci.

C: Comments on Technologies

There are a number of problems associated with identifying, cloning and characterizing microsatellites. These usually concern: a) Identifying repeats from a targeted chromosomal/genomic area; b) Accessing repeat flanking DNA sequences from which to design PCR primers; c) Poor amplification of alleles by PCR; and d) Visuallizing alleles amplified by PCR and correctly determining the genotypes of individuals.

Approaches to enriching for sequences from a targeted chromosomal region include screening libraries constructed from flow sorted chromosomes, somatic cell hybrids, and microdissected chromosomes. Each method has it's disadvantages. Flow sorted chromosome specific libraries are often contaminated with sequences from other chromosomes. Libraries constructed from somatic cell hybrids (a radiation hybrid for example) can often enrich for a very specific region of a human chromosome. Though these often cannot be fully characterized cytogenetically for their human complement of chromosomes due to breakage and rearrangments, the main difficulty is isolating human sequences away from the background of rodent clones. Since CA and non CA repeats are abundant in mammals, usually human clones must be identified by screening the library with total human DNA. A drawback to hybridizations with total human genomic DNA is that clones which contain only single copy human sequences may not be identified by this method (Neve *et al.*, 1986). The construction of libraries from microdissected chromosomes is also very specific for a targeted region. The predominant limiting factor

for success, as defined by the average insert size and the number of clones in the library, has been the very small amounts of DNA that can be isolated from 20-150 "scrapes" of metaphase chromosome spreads. The "micro-cloning" of picogram or sub-picogram quantities of DNA, requires microscopic manipulations in nanoliter droplets for most of the subcloning steps, such as deproteinization, restriction digests, ligations and phenol extractions (Kao, 1990). The microdissection of the Langer-Giedion syndrome region of 8q by Ludecke *et al.* (1989) yielded after PCR amplification a library of 20,000 clones with inserts that ranged from 52-350 bp (mean= 150 bp). Of 50 clones characterized, 80 % were single copy, non-repetitive sequences. In addition, 50 % of these clones detected deletions in two LGS patients, demonstrating the specificity of the dissection technique. Another report by Lichter *et al.* (1990) demonstrated the feasability of using PCR primers derived from ubiquitous human repeats (*Alu* and *LINE1*) to amplify DNA sequences between various orientations of repeats. An advantage to this procedure is that nanoliter droplet cloning manipulations are avoided since the cloning steps occur after the amplification of DNA by PCR. Lichter speculates that due to the relatively close spacing of the human specific repeats (approximately 4000 bp between *Alu's*), the PCR product and therefore the library, has a good chance of being highly representative of the DNA sequences present in the dissected region. The specificity of the amplification of human sequences using *Alu/LINE* derived PCR primers, was demonstrated by reverse *in situ* hybridizations of amplified human-rodent hybrid cell genomic DNA to human chromosomes. Although these results are encouraging, the construction of libraries from microdissected chromosomes remains technically demanding and generally unavailable to most researchers. (L. Allen, pers. comm.)

Screening flow sorted libraries:

To enrich for sequences from chromosome 21 the following approaches were used to identify clones from a flow sorted chromosome 21 library:

1) The Lehrach flow sorted chromosome 21 library was screened with previously mapped single copy probes. The clones identified were screened for the presence of STRs. An advantage to this approach is that if a locus of interest does not have a good polymorphism associated with it, there is a chance to increase it's informativeness by finding an associated microsatellite.

2) Lehrach's library was screened with STR repeat probes. The positive clones were subcloned and characterized without further mapping. This method has the potential for finding microsatellites which extend or fill gaps in a linkage map.

3) Cosmids from P. de Jong's flow sorted library which were known to lie in areas of interest by *in situ* hybridization were screened for the presence of STRs.

All methods were applied with varying degrees of success.

Table 9 summerizes the total number of cosmids which were screened, the number of repeat positive clones identified, and the number subcloned, sequenced, amplified and developed into markers.

TABLE 9: Total numbers

# of cosmids screened for CA or tetramer repeats: 227
# of positive clones: 71
# of cosmids subcloned: 29
# of subclones sequenced: 63
# of repeats found: 27 repeats from 23 cosmids
# of repeats PCR amplified: 21
# of polymorphic repeats with ≥4 alleles: 15
# of polymorphic repeats with heterozygosities ≥0.70: 10
# of repeats regionally mapped: 16*
# of polymorphic repeats on chromosome 21: 8
# of genotypes determined for all markers= approximately 4000
# of repeats mapped in 40 CEPH families: 9

* includes D11S236, a nonpolymorphic sequence tagged site

Results: A total of 227 cosmids from flow sorted chromosome 21 libraries were screened for the presence of repeats: 18 cosmids were identified with single copy probes and 17 by *in situ* hybridization to chromosome 21 with the remainder being unmapped clones. A total of 70 repeat positive clones were identified: 8 each from cosmids identified with single copy probes and in situ hybridization to chromosome 21; 55 were from unmapped clones. Twenty nine cosmids or phage were subcloned; 63 subclones were DNA sequenced; 27 repeats were found from 23 cosmids of which 21 were amplified by PCR. Fifteen polymorphisms have been regionally mapped and published; 10 microsatellites have heterozygosities ≥0.70.

Eight microsatellites are from chromosome 21. Of 8 previously unmapped CA positive clones characterized from Lehrach's library, 3 map to chromosome 21 (D21S168/S172/S198). Four polymorphisms were cloned from cosmids identified with single copy probes: Two map correctly (D21S11, D21S13); D21S167 maps to 21q22.2 instead of 21q11.2 and D4S233 to the wrong chromosome. D21S218 and D21S236 are from cosmids which were mapped prior to screening by in situ hybridization to human chromosome 21.

Discussion: The abundance of CA repeats and other STRs has been discussed. In theory, finding a repeat from a region of interest should not be a problem. Of main concern are the available resources which define the target region. In this dissertation we were provided with chromosome 21 flow sorted cosmid libraries. Lehrach's library is described to have 3-4 genomic equivalents of chromosome 21 with 70% of cosmids deriving from this chromosome (Nizetic *et al.*, 1991). As a result, a strategy of characterizing randomly selected CA repeat containing cosmids was pursued with the expectation that a majority of these would derive from chromosome 21. The primary objective was to find cosmids which were not from any previously defined locus and which would fill gaps on the linkage map of chromosome 21. In practice this approach was of marginal utility and quite

dependent on the representation of chromosome 21 sequences in the library. The most disappointing result is that of 8 previously unmapped CA positive clones characterized from Lehrach's library, only 3 mapped to chromosome 21 (D21S168/S172/S198). D21S172 is a high quality marker near D21S13 in 21q11.2, but overall, the yield of chromosome 21 microsatellites was poor. Our conclusion that less than 70% of cosmids in this library were from chromosome 21 was confirmed by Antonarakis, who found it necessary to construct a new flow sorted chromosome 21 library from which to isolate STRs (Antonrakis pers. comm.). Chromosome 21's small size allows it to be separated from large chromosomes by fluoresence-activated flow sorting of metaphase chromosomes (Tantravahi *et al.*,1988). This also means that fragments from larger (broken) chromosomes can sort with chromosome 21, and be incorporated into the library as contaminants. Contamination with other chromosome sequences in flow sorted libraries is not uncommon, as again evidenced by Antonarakis who found a number of repeats which mapped to other chromosomes in his new library. In our laboratory, a number of repeats developed from a flow sorted X-specific library mapped to other chromosomes (Luty pers. comm.). A humorous occurrence at a meeting on chromosome 21 at HGM 11 was a researcher's claim to have mapped a gene to chromosome 21 by virtue of it's being isolated from a flow sorted library!

Cosmids identified with single copy probes were screened for repeats to improve the PIC of 4 loci intimately connected with FAD (Table 10). These were D21S1, D21S11, D21S13 and D21S16. D21S1 and D21S11 define the distal end of the region of 21q in which we are most interested; D21S13/D21S16 define the proximal end (see Figure 4). Haplotypes must be constructed between D21S1/S11 and D21S13/S16 in order to improve the PIC of these loci. One good polymorphic microsatellite from each region would eliminate the need for haplotyping. Twelve cosmids were identified with the locus specific probes and all four probes identified a subset of cosmids which contained repeats.

Table 10 lists the single copy probes which were used to identify cosmids from specific loci, the PICs for the associated RFLPs, the number of repeats identified from the cosmids and the resulting microsatellite polymorphisms.

Table 10: Chromosome 21 single copy probes used to screen Lehrach's 96x96 arrayed library.

Locus	probe	RFLP PIC	Location	# of STR*	Result
D21S1	228C	0.36	q21.1	2 CA	D4S233
D21S11	236B	0.34	q21.1	2 CA; 1 AGAT	D21S11
D21S13	G21RK	0.36	q11.2	3 CA	D21S13
D21S16	NH2	0.20	q11.2	1 CA	D21S167
D21S8	245D	0.37	q21.2	0	

* # of STRs within cosmids identified with the single copy probe

Four polymorphisms were cloned- 3 CA repeats and an AGAT repeat. Two mapped correctly (D21S11, D21S13). D21S167 was isolated from a cosmid identified with D21S16 by Dr. Goate and maps to 21q22.2 instead of 21q11.2 . The D21S11 tetramer and D4S233 were cloned from cosmid #17 which was identified with a single copy probe for D21S11. This contradiction appears to result from the occurence of a mixture of cosmids in a single well of the microtiter plates used to store the cosmids from which subclones were made. This has subsequently been described by Lehrach and provides the most likely explanation for our result. PCR from single bacterial colonies and Southern blots of relevant cosmids show that D4S233 is derived from cosmid A07119 identified (incorrectly) with D21S1. On the basis of our results, microsatellites can be successfully developed from clones identified by locus specific probes. However, these require additional characterization by linkage analysis, *in situ* hybridization or somatic cell mapping panels before their map position can be assumed to be correct.

The results from the previous two screenings were 6 microsatellites on chromosome 21 (5 of index quality) which clustered into two groups: one group of 3 proximal to q21.1 (or D21S11); and one group of 3 distal to q22.2 (or D21S167). The gap between these

was substantial and later shown to be >30 cM. In order to avoid the attendent problems associated with characterizing unmapped clones and identifying cosmids with single copy probes, a third approach was employed— cosmids previously mapped to chromosome 21 by in situ hybridization were screened for the presence of repeats (see Table 11). Eight of 17 cosmids from de Jong's library were CA positive. Of these only 4 were specific to chromosome 21 since the other cosmids contained repetive sequences homologous to other chromosomes. Two cosmids in particular were characterized: 21A5 (D21S236) and 21B2 (D21S218). D21S218 mapped to q22.1 and was developed in order to close the gap between the distal and proximal cluster of microsatellites previously mentioned, allowing our microsatellite map to be constructed. D21S236 mapped to q11.2 and was developed with the hope of finding a marker between the centromere and D21S13. To increase the potential for developing useful microsatellites from the limited set of clones identified with single copy probes and *in situ* hybridization to chromosome 21, colony lifts of subclones which had been screened with CA repeat probes were rescreened with a pool consisting of 3 tetramers: (AAAT)n, (AAAG)n, & (AGAT)n. These repeats are associated with *Alu* sequences in the human genome and are reported to be abundant and polymorphic (Edwards *et al.*, 1991) Subclones from two of 7 cosmids screened with the tetramers contained a repeat by Southern blot analysis. This screen resulted in the development of the highly informative D21S11 tetramer.

The D11S35 and D11S286 microsatellites were successfully developed from previously mapped phage which identify RFLPs (Maslen, 1987). D11S35 is an index marker on chromosome 11 and was the first independently cloned and published CA microsatellite from our laboratory. D11S286 was the first non polymorphic CA repeat we characterized; 8 others have now been cloned, confirming that monomorphic microsatellite loci are not rare or unusual. Though monomorphic, D11S286 is a "sequence tagged site" (STS) since it's position is well defined on chromosome11. Sequence tagged sites facilitate the building of

Table 11 lists cosmids provided to us by Dr. Pieter de Jong. These are from a flow sorted chromosome 21 cosmid library and were *in situ* hybridized to human chromosomes at the Lawrence Livermore National Laboratory. The chromosomes which the cosmids hybridized to are noted.

Table 11: De Jong cosmids previously mapped to chromosome 21 by in situ hybridization.

CLONE	MAP LOCATION ON CHROMOSOME 21	OTHER CHROMOSOMAL LOCATIONS
21A5 J52-O	21q11.2	
21A11 J52-O	21q21.3 or q22.1	
21B2 J52-O	21q22.1	
21B4 J52-O	21p13	13p13, 14p13, 15p13, 22p13
21B6 J52-O	21q21.3 or q22.1	
21B8 J52-O	21q22.1	4p16, 7p22, 7q36, 10p13, 12q13, 13q14
21B10 J52-O	21p13	13p13, 14p13, 15p13, 15p12, 22p13
21B12 J52-O	21p13	13p13, 14p13, 15p13, 15p12, 22p13
21C11 J52-O	21q21.3 or q22.11	
21D8 J52-O	21cen, 21p13	4q35, 9cen, 13p13, 13cen, 14cen 15p13, 15cen, 20cen, 22p13, 22cen
21D10 J52-O	21q11	
21E1 J52-O	21cen, 21p13	4q35, 9cen, 13cen, 14cen, 15cen, 15p13, 16cen, 16q24, 19cen, 20cen, 22cen
21E5 J52-O	21q22.1	
21E11 J52-O	21q22.2	
21F2 J52-O	21q22	
21F4 J52-O	21q22.2 or q22.3	
21F6 J52-O	21cen, 21p12	13cen, 14cen, 15cen, 22cen

Clones mapping to multiple p13 locations (such as 21B4) probably contain satellite DNA.

contiguous physical maps by providing markers which use PCR to identify YACs, cosmids or pulse field restriction enzyme fragments which contain the locus.

Though a disappointing number of microsatellite polymorphisms we developed did not map to chromosome 21, they have proved to be quite valuable as markers on other chromosomes. In particular, D14S43, D9S55, D18S37 and D11S35 have been designated reference markers and are currently being incorporated into index maps under construction for the US human genome initiative. The genotypes for 40 CEPH families were determined for D14S43 and D18S37 and sent to CEPH for inclusion in their data base. A number of polymorphisms are being used by researchers worldwide for human disease studies as evidenced by journal citations and the requests for locus information which we receive periodically .

Rodent-human somatic cell panels have been important for mapping loci to specific chromosomes. Their use became necessary in my work when I realized that many of the unmapped cosmids from Lehrach's library did not originate from chromosome 21. Typically, microsatellites from cosmids identified with single copy probes or *in situ* hybridization were typed in a number of CEPH families; linkage analysis followed using markers from the region expected to contain the microsatellite. If linkage was not detected, the Bruns and Mohandas cell hybrid panels for all 24 human chromosomes were used (Bufton *et al.*,1986; Mohandas *et al.*,1986). The human chromosomes contained in each cell line of a mapping panel have been characterized by cytogenetic methods and/or the presence of expressed sequences or DNA probes. The microsatellite is typed through the panel and each cell line is scored for the presence or absence of the repeat. Discordant results are common when using hybrid cell mapping panels because the cell lines may not be stable. In addition, when the complement of human chromosomes contained in a cell line is being characterized, chromosomes contained in a small percentage of cells, chromosome fragments or rearranged chromosomes might not be detected. On the other hand, PCR can

detect a sequence which can be not be confirmed to be present cytogenetically or by DNA probes resulting in an apparent discordancy. Figure 10 shows the Bruns mapping panel and the scoring for D14S43. Linkage analysis provided the final confirmation of the map position of the microsatellites. In no case were there discrepencies between the results obtained by linkage analysis and the mapping panels; nor were there loci which could not be mapped to a specific chromosome when the Mohandas and Bruns panel were both used. Table 12 lists the method used to identify the cosmid from which the microsatellites were cloned and the method(s) that was required to confirm their chromosome of origin.

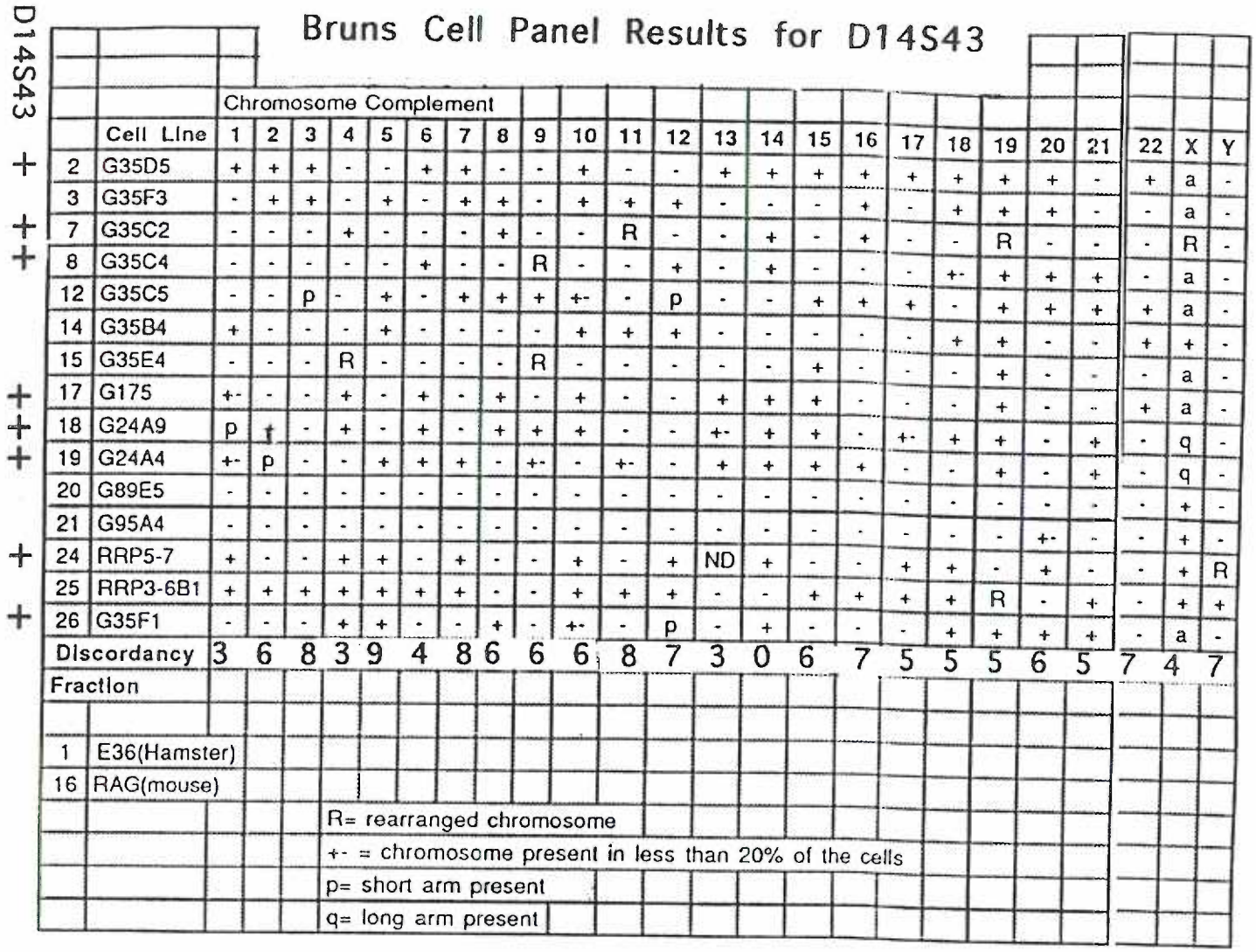
Table 12: Methods used to identify cosmid clones and confirm their chromosome of origin
a: Source: R= unmapped cosmid; I= in situ mapped; S= single copy probe
b: Confirmation: I =in situ hybridization L=linkage C= hybrid cell panel

Locus	Source	Confirmation
D21S236	I	L
D21S172	R	I/L
D21S13	S	L
D21S11	S	L
D21S218	I	L
D21S167	S	C/L
D21S168	R	C/L
D21S198	R	C/L
D11S35	S/I	L
D1S117	R	C/L
D9S55	R	I/L
D14S43	R	I/C/L
D4S233	S	C/L
D18S37	R	C/L
D1S116	R	I/C/L

Figure 10: Chromosome contents of the Brun's mapping panel hybrid cell lines. The designations are: (+) presence or (-) absence of a human chromosome; (R) rearranged chromosome as determined by the disruption of a syntenic group or cytogenetic abnormality; (±) chromosome present in less than 15% of cells; (p) presence of a marker for the short arm and absence of a marker for the long arm; (q) presence of a marker for the long arm and absence of a marker for the short arm; (a) Xq24-qter. In counting discordencies, hybrids with a rearranged chromosome or those where the chromosome is present in less than 15% of the cells were excluded.

The column labeled D14S43 indicates the presence or absence of PCR amplification of a cell line with D14S43 The number of discordent signals for each line is shown; there were none for chromosome 14.

Bruns Cell Panel Results for D14S43



D14S43		Cell Line	Chromosome Complement 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
+	2	G35D5	+	+	+	-	-	+	+	-	-	+	-	-	+	+	+	+	+	+	+	+	-	+	a	-
+	3	G35F3	-	+	+	-	+	-	+	+	-	+	+	+	-	-	-	+	-	+	+	+	-	-	a	-
+	7	G35C2	-	-	-	+	-	-	-	+	-	-	R	-	-	+	-	+	-	-	R	-	-	-	R	-
+	8	G35C4	-	-	-	-	-	+	-	-	R	-	-	+	-	+	-	-	-	+-	+	+	+	-	a	-
	12	G35C5	-	-	p	-	+	-	+	+	+	+-	-	p	-	-	+	+	+	-	+	+	+	+	a	-
	14	G35B4	+	-	-	-	+	-	-	-	-	+	+	+	-	-	-	-	-	+	+	-	-	+	+	-
	15	G35E4	-	-	-	R	-	-	-	-	R	-	-	-	-	-	+	-	-	-	+	-	-	-	a	-
+	17	G175	+-	-	-	+	-	+	-	+	-	+	-	-	+	+	+	-	-	-	+	-	-	+	a	-
+	18	G24A9	p	f	-	+	-	+	-	+	+	+	-	-	+-	+	+	-	+-	+	+	-	+	-	q	-
+	19	G24A4	+-	p	-	-	+	+	+	-	+-	-	+-	-	+	+	+	+	-	-	+	-	+	-	q	-
	20	G89E5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
	21	G95A4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+-	-	-	+	-
+	24	RRP5-7	+	-	-	+	+	-	+	-	-	+	-	+	ND	+	-	-	+	+	-	+	-	-	+	R
+	25	RRP3-6B1	+	+	+	+	+	+	+	-	-	+	+	+	-	-	+	+	+	+	R	-	+	-	+	+
+	26	G35F1	-	-	-	+	+	-	-	+	-	+-	-	p	-	+	-	-	-	+	+	+	+	-	a	-
	Discordancy		3	6	8	3	9	4	8	6	6	6	8	7	3	0	6	7	5	5	5	6	5	7	4	7
	Fraction																									
	1	E36(Hamster)																								
	16	RAG(mouse)																								

R= rearranged chromosome

+- = chromosome present in less than 20% of the cells

p= short arm present

q= long arm present

In summary, 7 of 8 previously unmapped cosmids required the use of the cell hybrid panels in order to identify the chromosome from which they originated. Two of 4 microsatellites from cosmids identified by single copy probes required the use of the cell hybrid panels (D21S167, D4S233). The positions of the microsatellites which originated from cosmids/phage previously mapped by *in situ* hybridization were confirmed by linkage analysis without need of further characterization.

Subcloning CA repeats

Once repeat bearing cosmids were isolated, our first approach was to subclone Sau 3A restriction digests into the commercially available vector pTZ18U (US Biochemicals). Sau3A has a 4 base recognition sequence which is predicted to cut frequently leading to appropriately sized inserts for accessing repeat flanking DNA sequences from which to design PCR primers. Subcloning efficiency was increased by partially filling the vector and insert ends with the first two nucleotides of the recognition sequences. Compatible cloning ends between insert and vector are maintained while self ligations between vector and insert concatemerization are eliminated. (Y. Nakamura, pers comm.) This approach was successfully used to subclone D11S35 and D11S286. However it soon became apparent that many CA positive Sau3A subclones contained inserts that were too large to allow the repeat to be sequenced with vector-based sequencing primers. To avoid cloning large inserts, other restriction enzymes based on 4 base recognition sequences were examined. Alu I, Hae III and double digests with Alu I & Hae III were compared to Sau3A by Southern blotting cosmids to size the repeat containing fragments. This comparison revealed that Sau3A tended to generated larger inserts and a higher proportion of partial digests than the other enzymes.

Sizing repeat containing fragments prior to subcloning allows one to select the enzyme that does not result in excessively large or small fragments. Since one is not sure where within any particular restriction fragment a repeat may lie, the best general strategy is

to clone with an enzyme that will allow the CA containing insert to be sequenced completely using forward and reverse sequencing primers. This has worked with varying success, the most common problem being repeats which abut the vector cloning site resulting in little or no DNA sequence on one side of the repeat from which to design PCR primers. If one is characterizing unmapped cosmids, one can stop at this point and proceed to the next cosmid or subclone. However, if the repeat is known to be in an area of interest, it usually becomes necessary to resubclone the cosmid using different restriction enzymes in order to produce a different sized insert. One primer from a vector adjacent subclone can be designed and used as a probe to identify the proper new subclone, and as a sequencing primer to yield the desired flanking sequence. This strategy was used to characterize D21S172 .

Once one has subcloned and accessed repeat flanking DNA sequences additional problems may arise. These are seen as non-Mendelian inheritance of alleles, poor or failed amplifications and unscorable genotypes/alleles.

PCR

Conditions for the amplification of microsatellite polymorphisms by PCR are usually empirically determined. Factors which we have found to affect amplification include the source (manufacturer) of the polymerase, type of thermal cycler, PCR reagents/components, number, length and temperature of PCR cycles and primer concentration. PCR amplifications which produce little product usually result from poorly designed primers. In the set of microsatellites described here, poor amplifications resulted from the incorporation of vector sequence in a PCR primer, a 3' mismatch in one PCR primer due to a single nucleotide polymorphism and the use of short primers.

The primer pair developed from an original subclone of D1S116 worked poorly. Amplification could be acheived only with difficulty and extensive adjustment of annealing temperatures and PCR conditions. A closer examination of the DNA sequence revealed that 10 nts. of vector had been incorporated into the 5' end of a 20 nt. primer. In retrospect, it is

surprising that any PCR product could be formed with this primer. A new, correct primer was never designed because a different repeat was found in another subclone from the same cosmid. This repeat was fully characterized and published as D1S116.

The effect of a mismatch near the 3' end of a PCR primer and the importance of confirming Mendelian inheritance was demonstrated by D9S55. Characterization of 40 random individuals showed ≈50% heterozygosity for this locus. When two CEPH families were genotyped, it was observed that each family appeared to transmit a null allele. This was noted by the lack of transmission of a parental allele to offspring (Figure 11). The presence of a null allele in a single family, though interesting, could be presumed to be to an isolated event which prevents PCR. However in this case multiple families were involved. To explore the basis for the null allele, D. Browne developed new primers and amplified sequences from transmitters of the null allele and "normals". By sequencing several alleles, he found a single nucleotide polymorphism near the 3' end of one of the original primers which had previously prevented amplification of an allele (Figure 12).

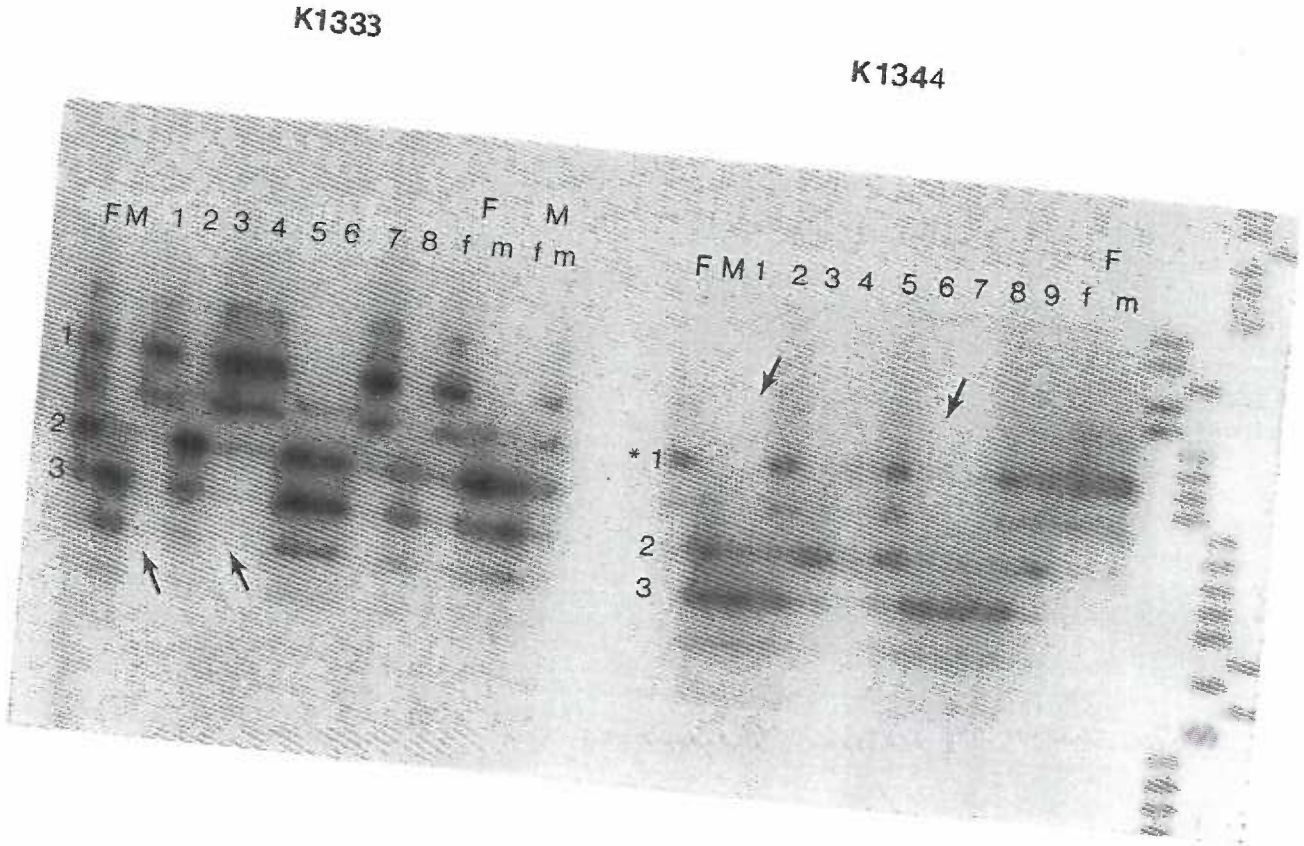
Figure 12: DNA sequence for D9S55 (subclone 1F6). Original primer sequences are underlined; the position of the nucleotide polymorphism which prevented amplification of an allele in Figure 11 is shown in bold type.

5'-
CTATTTCAACAGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
A

GTGTGTGTGAAACTGCTGCCTGGAATTTTCTCC-'3

Figure 11: Autoradiograph of the alleles in CEPH families K1333 & K1344 for locus D9S55. The lack of transmission of a parental allele to an offspring can be seen in numerous individuals in each family (arrows). In family K1333 children do not appear to receive an allele from mother (M) and in family K1344 from father (F); "f & m" are the grandparents.

D9S55

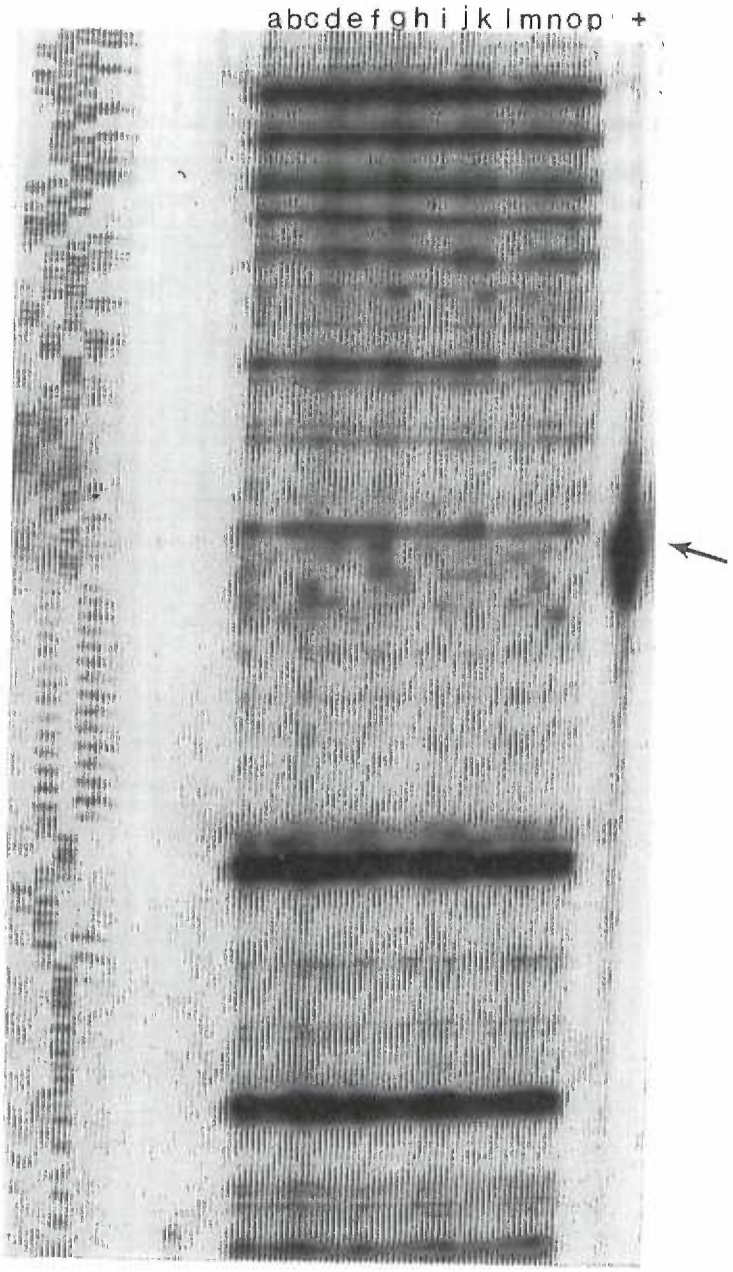


Development of a new primer pair allowed the characterization of the locus to be completed. The sensitivity of PCR to mismatches at the 3' end of primers was demonstrated as was the need for confirming that genetic markers are inherited in a Mendelian fashion. "Null" alleles will be masked in individuals who will appear to be homozygous for another allele, potentially confounding linkage analysis in small or incomplete families, where transmission of the null allele may not be detected.

The importance of length and annealing temperature of primers has been demonstrated for a number of loci. In particular, the amount of amplified product for D21S198 and D21S11 increased when new, longer primers were designed. The original primers for D21S198 were 17mers with a predicted annealing temperature of 54°C. When radioactive nucleotide was incorporated into the PCR reaction mix, this locus was barely discernable and completely unscorable (Figure 13). This was presumably due to PCRprimers annealing in a nonspecific manner to genomic sequences. The alleles of a locus can be obscured if they co-migrate with a nonspecifically amplified product. The competition for PCR reagents between nonspecific products and legitimate target can prevent strong amplification of the locus. It is important to note that once a nonspecifically primed product is produced it serves as a template in subsequent PCR cycles and becomes amplified. In Fig.13 the area where the proper PCR product is expected can be determined by the position of the positive control (arrow)- in this case, the original cosmid from which the microsatellite was cloned. Amplification of this cosmid does not result in the extra bands observed in genomic DNA, which confirms that sequences unrelated to the locus are being amplified by the primers. A typical solution to this problem is to increase the annealing temperature of the primers which increases the specificity of the PCR by discriminating against heteroduplexes formed between mismatched primers and target sequences. These have lower melting temperatures (are less stable) than properly annealed primers. With D21S198, 17mers proved to be (too) sensitive to small increases in the annealing temperature such that PCR would fail, presumbly due to the lack of annealing of

Figure 13: Internally labeled PCR products of16 CEPH parents for locus D21S198 resolved on a 6% DNA sequencing gel showing the presence of nonspecifically amplified constant bands. The expected position of the alleles for the locus can be determined from the position of the positive control (arrow); they are unscorable.

D21S198



primers. The development of "blotting and probing" protocols were necessary to visualize the alleles at this locus when using the original primers. (discussed later) New primers 22 & 23 nucleotides long were made with substantial improvements in the amount of amplified product produced by PCR.

With D21S11 a number of primer pairs were developed. All amplified but with varying degrees of reliability. D21S11 is a very long compound tetramer repeat which was very difficult to sequence. As a result, the possibility existed that one primer might have missed or incorporated an incorrect base, perhaps accounting for some of the PCR failures which plagued this locus. Four bases were added to the 5' end of each primer of the primer set which had been the most reliable for PCR. No changes were made at the 3' end in order to avoid worsening the situation by possibly introducing an incorrect base at the end of the primer which is critical for allowing extension in PCR. This simple change increased the amount of product and reliability of PCR for the locus.

We have adopted some general rules for primer design as a result of our experiences. Primarily, they are that longer primers (>22 nts.) of medium GC content with target annealing temperatures of 60-65°C are likely to amplify best. We feel that these provide the necessary locus specificity and yet are forgiving by having a broad range of acceptable PCR conditions. This is important to minimize the time required to characterize a locus, since it has been demonstrated that poor primers can be made to work with considerable effort (D1S116). Our data which suggests that longer primers are more reliable is based on experiences with D18S37, D21S218, D21S198 and D21S11.

In summary, there are a number of ways in which primers determine the efficiency of PCR amplifications. These are dictated primarily by the length and annealing temperature of the primers used in PCR. Longer primer sequences are more unique in the genome and more stable at higher temperatures than short primers. Equally clear is that only the 3' end of a primer needs to anneal to a template in order for there to be extension and amplification. Bands (products) produced from non specific priming in PCR are

reproducible and "constant". Once these templates are created they amplify efficiently and compete with the desired locus for polymerase and reagents. Wu *et al.* have studied the effect of temperature and oligonucleotide primer length on the specificity and efficiency of amplification by polmerase chain reaction (Wu *et al.*, 1991). They propose that "the successful priming of an oligonucleotide on a DNA template is governed by two variables: the rate of primer dissociation from the primer-template complex before initiation of polymerization and the rate at which the DNA polymerase extends the primer until a stable primer-template complex is formed. PCR is governed kinetically; once a transient association between primer and template DNA has occurred, the addition of the first few nucleotides to the primer allows the continued extension of the primer until the product is complete" (Wu *et al.*, 1991). The temperature of polymerization (Tp) is defined as the highest temperature which allows extension and amplification of a sequence by PCR. They note that the Tp is generally 5-10°C higher than the melting temperature (Tm) of the primer, which is defined as the temperature at which 50% of the oligonucleotide duplex dissociates. "Although the temperature Tp is greater than Tm, where primers are not expected to anneal stably to the template, priming occurs by the elongation of the primer when it interacts transiently with the template at the annealing site. In summary, whether a primer is to be extended or not depends on which event will predominate--primer dissociation or primer elongation. At higher temperatures, shorter oligonucleotides will dissociate more rapidly compared to longer oligonucleotides" (Wu *et al.*, 1991).

The observation of Wu *et al.* provides a satisfactory explanation for the increased amount of desired product we observe when using "Touchdown" PCR protocols (Don *et al.*, 1991). In touchdown PCR, the annealing temperature for the first cycle is 10°C higher than the "target" or empirically determined annealing temperature for the primers (essentially the Tm). The temperature is reduced 1°C per cycle until the target annealing temperature is reached, after which an additional 20 PCR cycles are performed.

Allele Resolution:

There have been a number of technical improvements which have eased the difficulties associated with accurately determining the genotypes of individuals when using microsatellite polymorphisms. These are the use of sequencing gels containing 32% formamide, the transferring of unlabeled fragments to membranes, and the hybridization of repeat sequence (or locus specific) oligomer probes to the transferred fragments in order to visualize them.

Typically, PCR products are resolved on standard denaturing polyacrylamide DNA sequencing gels in order to determine their length in nucleotides. Such gels are capable of resolving the alleles for many loci. However, some loci give rise to smears (or ladders) which prevent the accurate determination of genotypes. The alleles of dinucleotide repeats usually vary in increments of two nucleotides or one repeat unit. (Weber & May, 1989; Edwards *et al.*, 1991; Litt & Luty, 1989). There are a number of problems associated with scoring alleles which are inherent to CA microsatellites. These are the presence of slipped strand mispaired products, non specific PCR products, and the potential for sequences to form secondary structures. Their resolution concerns the denaturing capacity of gels, the specificity of PCR amplifications and the method of radioactively labeling alleles.

Multiple bands generated in PCR amplifications (shadow bands) are a particular characteristic of dinucleotide repeats. They can be the result of:

1) Slipped strand mispaired products generated during PCR which present as shorter bands at two nucleotide (one repeat) increments. Often they can interfere with the scoring of alleles which vary by 1 or 2 repeat units (Fig. 14B). Usually a locus will have a characteristic pattern of shadow bands based on their intensity and number which can help determine the genotype of an individual. Family relationships can also help to accurately determine a genotype.

2) Incorporating radioactive nucleotide into the PCR labels both strands in proportion to the frequency of that particular nucleotide. The complementary strands of DNA will have slightly different mobilities due to different base composiitons and sequences. If both strands are equally labeled and have different mobilities, each allele will produce two bands of equal intensity. Unlike slipped strand mispaired products, these usually do not differ from each other in nucleotide increments (Figure 14A).

A method of blotting PCR products to membranes directly from DNA sequencing gels was demonstrated by Luty based on a multiplex DNA sequencing method first described by Church and Kiefer-Higgins, (1988). PCR products are visualized with repeat or locus specific probes. Figures 14 A & B are of the same individuals amplified under indentical PCR conditions except that Figure 14B shows unlabeled PCR products transferred to a membrane and probed with an end labeled PCR primer. When using locus specific probes (such as a PCR primer) only one strand of DNA will be visualized on the blot. This reduces the interference from internally label PCR products and facilitates the scoring of autoradiographs. This method does not reduce the interference due to slipped strand mispaired products.

When repeat sequence probes are used on membranes, nonspecific PCR products will not be seen since these are unlikely to contain a repeat. For example, D21S198 was unscorable when the PCR products were labeled by the incorporation of radioactive nucleotide during amplification (see previous Figure 13). When unlabeled product was generated under identical PCR conditions, blotted to membranes and probed with the $(CA)_{15}$ oligomer, the interfering background of labeled, nonspecific product was completely eliminated allowing the alleles at the locus to be seen and easily sized (Figure 15).

Figure 14: Panel A shows the internally labeled PCR products of 17 CEPH parents for locus D21S172. Panel B shows unlabeled PCR products from the same individuals transferred to nylon membranes and probed with a kinased PCR primer. The arrow points to a heterozygous individual whose genotype can be determined in panel B since only a single DNA strand is visualized by this method. Alleles in panel B show the characteristic microsatellite slipped strand mispaired PCR product.

D21S172

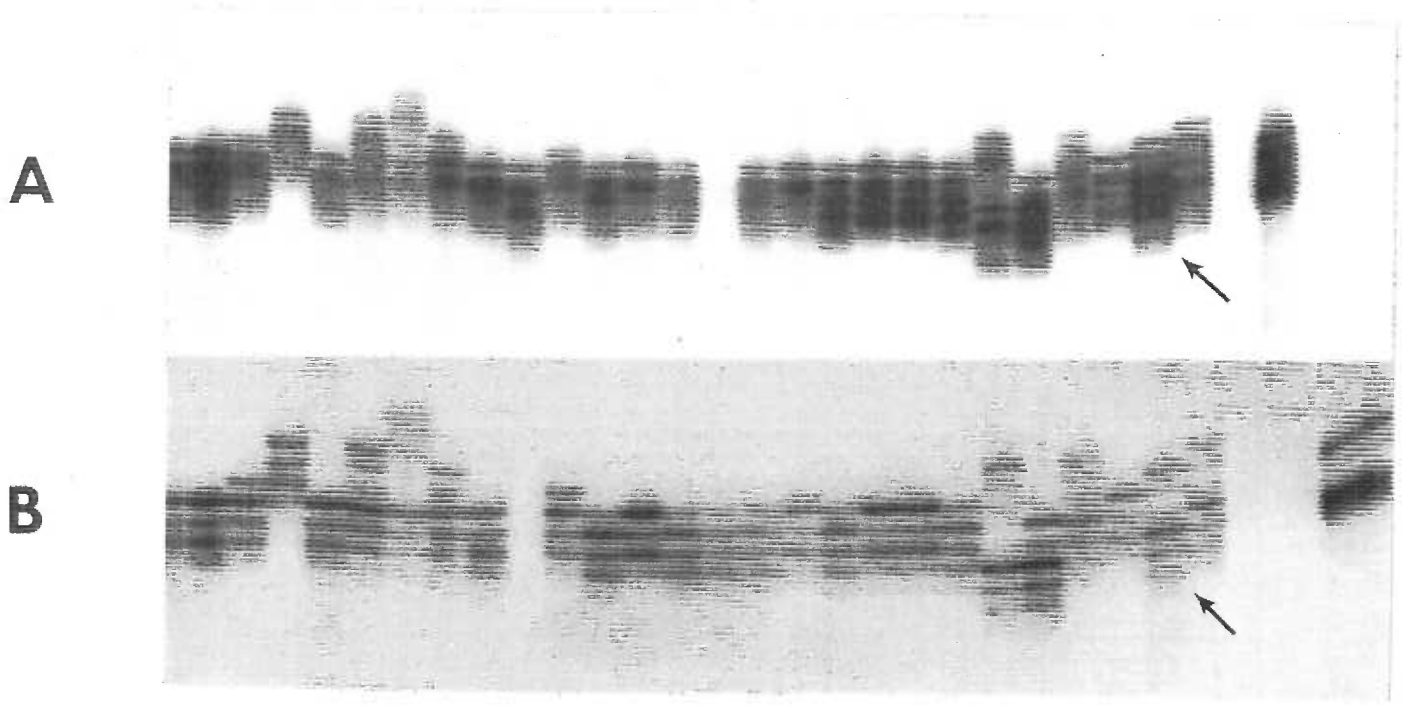


Figure 15: The unlabeled PCR products of 22 CEPH parents for locus D21S198 resolved on a 6% DNA sequencing gel, blotted and probed with kinased $(CA)_{15}$ oligomer. The annealing temeperatures and PCR cycles times were identical to that shown in figure 13; nonspecifically amplified product is not seen by this method.

D21S198

Smears are the result of incomplete denaturing of DNA and secondary interactions (conformations) of DNA formed between species present in the PCR. Unlike the ladders which form when the products of DNA sequencing reactions are resolved on gels, PCR results in a high concentration of few products of specific length. These may not resolve from each other due to the insufficient denaturing capacity of the sequencing gels. The ability to form secondary structures or remain undenatured is a function of the GC content, length, and sequence of a particular locus. For example, D14S43 would present as a smear on sequencing gels (Figure 16). This locus contains a GT repeat which has been duplicated in tandem along with flanking sequences (Figure 17A). The high GC content and sequence complexity of D14S43 clearly demonstrates a strong potential to form unusual secondary structures between complementary strands. Sometimes running a gel at higher temperatures (>55°C), and increasing the volume and formamide concentration of the loading dye added to the samples will help resolve the alleles of a locus. This approach was necessary for determining the genotypes for D14S43 in all 40 CEPH families.

If running a gel at higher temperatures will not resolve the alleles of a particular locus, the denaturing capacity of sequencing gels can be dramatically improved by adding formamide to the gel (Michel Georges, pers. comm.). Figure 18A and 18B show the same amplification products loaded on a normal and a formamide sequencing gel respectively. Figure 18A is virtually unscorable; the alleles in Figure 18B are fully resolved. The increased denaturing capacity of formamide sequencing gels has nearly eliminated all problems of allele visualization which involve secondary structures formed between PCR products.

Scale up:

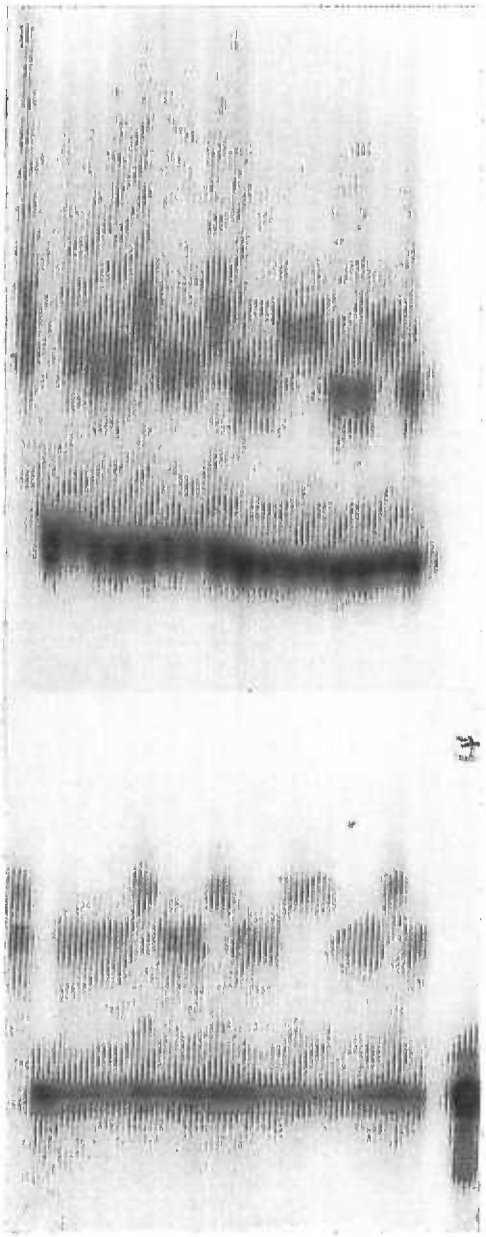
Our ability to determine the genotypes of large numbers of individuals has improved considerably as a result of the procedures described above. Newer PCR machines use 96 well formats and have rapid cycle times allowing hundreds of amplifications to be

performed in a single day. The use of 96 well microtiter plates and multichannel pipettors has increased the speed with which one can handle large numbers of samples. The ability to eliminate the interference of nonspecific PCR products by blotting and probing membranes allows gels to be loaded multiple times (i.e. multiple ladders). By using multichannel pipettors and sharkstooth combs which form 96 wells in a gel, over 300 samples can be loaded easily and quickly onto a single gel (Figure 19). For example, when D14S43 was typed through 40 CEPH families it took about 2 months to accomplish. Each individual required a separate sample tube, the PCR machine handled a maximum of 60 samples and averaged 3-4 hours per PCR run. One year later, 4 loci were typed through 40 CEPH families in a similar two month period.

When using multichannel pipettors, blocks of 24 samples can be loaded in two orientations (see previous figure 5). Any sample in one orientation does not share the same "neighbors" in the other orientation. This allows genotypes to be independently confirmed by loading the samples twice. For example, errors due to leaky wells or sample "spillage" in one orientation would be detected in the other orientation. Also, this format allows one to randomize samples ("break-up" families). This approach was used in determining the genotypes of the FAD families discussed earlier. These kindred were small, incomplete (no parents etc.) and difficult to DNA type by PCR. When the data was reassembled into "families", the general agreement in the alleles which were shared between related individuals considerably increased our confidence in the correctness of the genotypes.

Figure 16: Internally labeled PCR product of CEPH family K1362 for locus D14S43. Panel A and B show the same PCR product on a 5% DNA sequencing gel run "hot" at 95 watts (panel B) compared to a gel run at 70 watts (panel A).

D14S43: CEPH K1362



A

B

Figure 17A: The DNA sequence for D14S43 (subclone 2E12b). Primer sequences are underlined. A CA repeat with flanking sequences (shown in bold and outline type) has been tandemly duplicated.

Figure 17B: The sliding base composition (%G+C vs nucleotide position) for D14S43.

A

D14S43 (2E12b) DNA Sequence:

5'-

CCTGGAACACTCAGGCGAGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTTGGGAA

TGGGGCTGGGGGGGAGACAGGGTTGTGTGTGTGTGTGTGTGTGTGCGTGTTGGGAAT

GGGGCTGGGGGGGAGACAGGGTTGAGTGACCCAAAGTAGAAAGTGGCTCTGGATC

-'3

B

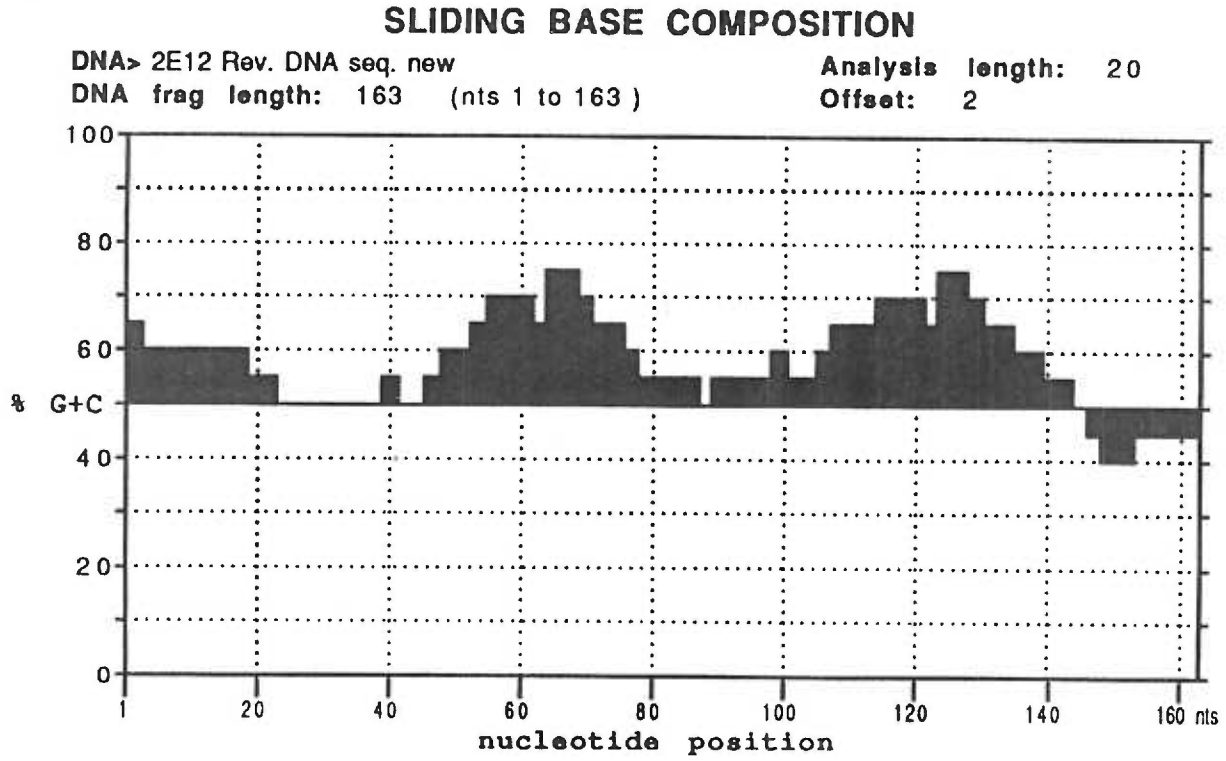
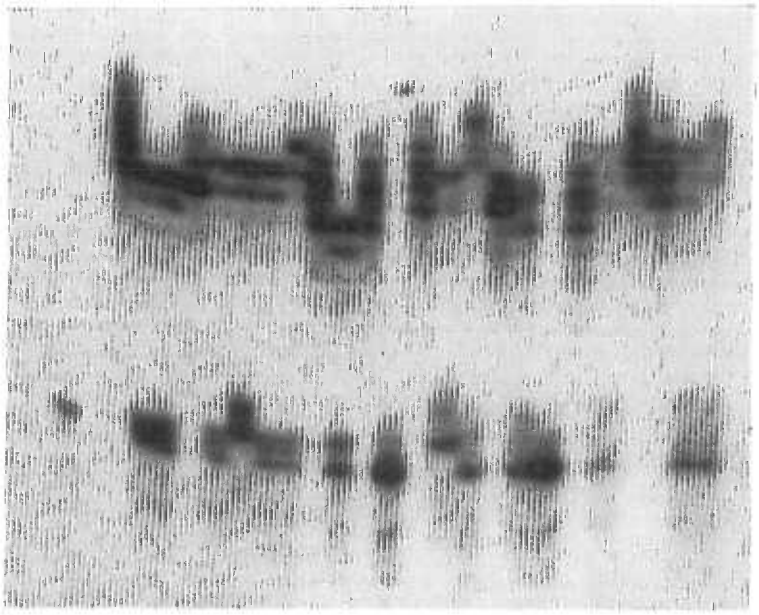


Figure 18: PCR products from 24 CEPH parents for locus D21S236 were blotted and probed with a (CA)15 oligomer. In panel A, the alleles were resolved on a standard DNA sequencing gel; in panel B, the same PCR products were resolved on a DNA sequencing gel containing 32% formamide.

D21S236

A



B

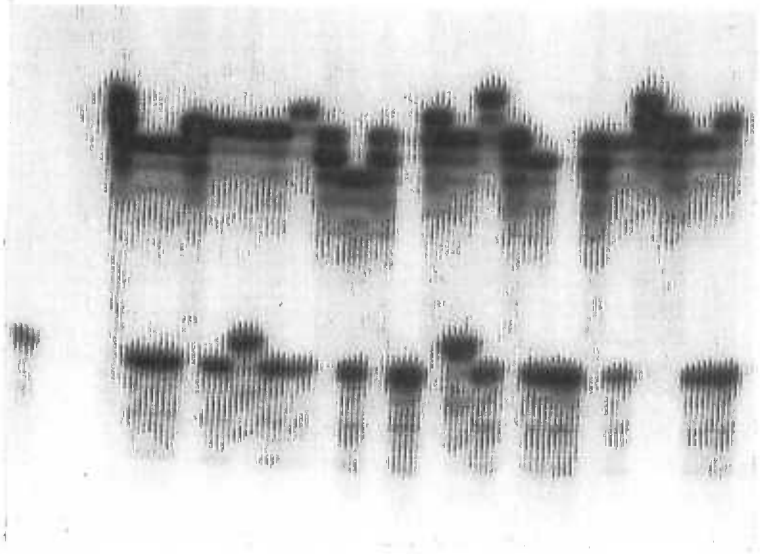
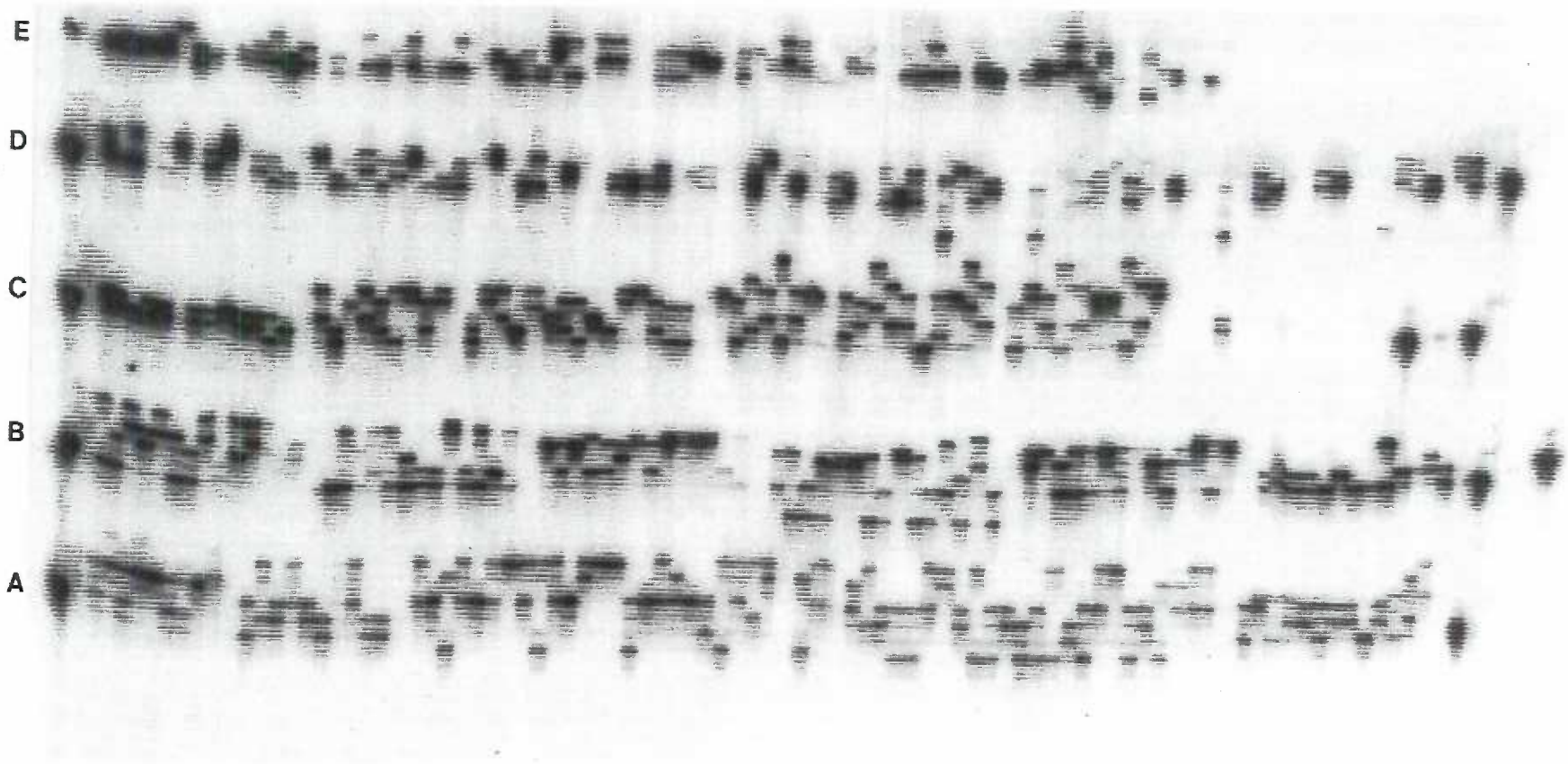


Figure 19: The genotypes of approximately 30 CEPH families for locus D21S11 are resolved on a single gel loaded 5 times (A-E) with a multichannel pipettor (>300 individuals).

D21S11



V: SUMMARY

In conclusion, short tandem repeats have revolutionized linkage map constructions by providing a polymorphism which is abundant, well distributed and highly informative in linkage studies. In the few short years since microsatellites were first described, (and this dissertation begun) well resolved linkage maps of all human chromosomes are begining to appear. We have demonstrated their value by cloning 16 repeats, 10 of which have heterozygosities over 0.70 and deriving a set of 8 polymorphisms from chromosome 21 which have been used to constuct linkage maps of this chromosome with much greater ease and resolution than maps constructed by less informative polymorphisms such as RFLPs. . Equally important, technology improvements have been developed which increase the efficiency and reliability of characterizing dinucleotide repeats.

Summary of CA Repeat Technology Applications:

1) CA repeats are abundant and informative enough to serve as the primary source of polymorphisms for genetic maps constructions.
2) Screening clones from mapped sources is the most efficient way of targeting specific chromosome regions unless the primary DNA source used in making a library is very well characterized.
3) Screening primary clones and subclones with the CA/GT copolymer probe will identify longer repeats on average than the $(CA)_{15}$ oligomer. Repeats so identified are more likely to have high PIC's.
4) The ease of characterizing a CA repeat by PCR can be enhanced by increasing the specificity of primers annealing to target. This is most readily accomplished by using "Touchdown" PCR protocols, computer designed primers, primers which are 22-26 nucleotides long and which have annealing temperatures >60°C.

5) Blotting unlabeled PCR amplified products to membranes and visualizing them with internal oligo probes has numerous advantages since only PCR products which amplify across the desired segment will be visualized.

   a) Non specific PCR products produced with kinased PCR primers or by the incorporation of radioactive nucleotide in amplifications are not seen because it is unlikely that these will contain a repeat and therefore do not hybridize to the probe.

   b) Locus specific products which result from the internal labeling of both DNA strands with radioactive nucleotide in PCR amplifications are not seen.

   c) The annealing temperature of PCR reactions can be below that required to specifically amplify radioactive PCR products (i.e less stringent). Reduced PCR annealing temperatures can increase the probability that all the samples in a PCR run will amplify. This is particulary important when temperatures vary with the position of the samples within a single PCR machine.

   d) Loading multiple ladders of PCR products on DNA sequencing gels becomes much more practical, since there will be no interfering bands from nonspecific amplification products.

   e) PCR samples which do not contain radioactivity are stable and can be stored indefinitely.

6) The increased denaturing capacity of formamide sequencing gels has virtually eliminated all problems of allele visualization which involve secondary structures formed between PCR products.

7) The use of 96 well format PCR machines and multichannel pippetors has greatly decreased (5-10 fold) the time it takes to set up amplifications by PCR. With the multichannel pippetor over 300 PCR reaction can be routinely loaded onto a gel quickly in two different orientations increasing accuracy of genotyping.

VI: REFERENCES

Antonorakis SE, Boehm CD, Giardina PJV, Kazazian HH. (1982) Nonrandom association of polymorphic restriction sites in the ßglobin gene cluster. Proc. Natl. Acad. Sci. USA 79:137-141.

Antonarakis SE, et al. (1991) Parental Origin of the extra chromosome in trisomy 21 as indicated by analysis of DNA polymorphisms. The New England Journal of Medicine 324: 872-6.

Antonorakis SE, Chakravarti A, McInnis MG, Sharma V, Avramopoulos D, Blaschak JE, Litt M, Blouin JL. (1992) Human Genome Map: Chromosome 21, part of Donnis-Keller *et al.* Science (submitted)

Armour JA, Wong Z, Wilson V, Royal NJ, Jeferies AJ. (1989)

Baran N, Lapidot A ,Manor H. (1991) Formation of DNA triplexes accounts fopr arrest of DNA synthesis at d(TC)n and d(GA)n tracts. Proc Nat Acad Sci USA 88: 507-511.

Bird TD, Sumi SM, Nochlin D, Schellenberg G, Nemens E. (1989). Familial Alzheimer's disease: A classification based on the analysis of 31 multigeneration pedigrees. Abstract in Alzheimer's Disease and Associated Disorders Vol 3 (1): pg 5.

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

Bufton L, Bruns GA, Magenis RE, Tomar D, Shaw D, Litt M. (1986) Am. J. Hum. Genet. 38: 447-460.

Burmeister M, Kim S, Price ER, de Lange T, Tantravahi U, Myers RM, and Cox DR(1991). A map of the distal region of the long arm of human chromosme 21 constructed by radiation hybrid mapping and pulsed field gel electrophoresis. Genomics 9:19-30.

Burmeister M, Cox DR, and Myers RM(1990). Dinucleotide repeat polymorphism located at D21S120. Nucleic Acids Res 18:4969.

Butler R, Ogilvie DJ, Elvin P, Riley JH, Finniear RS, Slynn G, Morten JEN, Markham AF, and Anand R (1992). Walking, Cloning, and Mapping with Yeast Artificial Chromosomes - A Contig Encompassing D21S13 and D21S16. Genomics 12:42-51.

Carritt B, Litt M. (1989). Report of the committee on the genetic constituition of chromosome 20 and 21. Cytogenet Cell Genet. 51:358-371.

CEPH version 5 data base (1991). LOD scores and recombination estimates.

Church GM Kieffer-Higgins S. (1988) Multiplex DNA Sequencing. Science 240:185-8.

Church GM, Gilbert W. (1991) Proc. Natl. Acad. Sci. USA 81, 1991

Collins FS. (1992) Positional cloning: Lets not call it reverse anymore. Nature Genetics 1: 3-6.

Conneally PA, Rivas ML. (1980) Linkage analysis in man. In Advances in Human Genetics, Vol 10:209-66, New York: Plenum.

Cotton RGH, Rodrigues NR, Campbell RD. (1988) Reactivity of cytosine and thymine in single base pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. Proc. Natl. Acad. Sci. USA 85: 4397-4401.

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.

Cox D R, and Shimizu N(1991). Report of the Committee on the Genetic Constitution of Chromosome-21. Cytogenetics and Cell Genetics 58:800-826.

Dausset J, Cann H, Cohen D, Lathrop GM, Lalouel JM, and White RL(1990). Program description: Centre d'Etude du Polymorphisme Humain (CEPH): Collaborative genetic mapping of the human genome. Genomics 6:575-578.

De Jong PJ, Yokobata K, Chen C, Lohman F, Pederson L, McNinch J, Van Dilla M. (1989). Human chromosome-specific partial digest libraries in lambda and cosmid vectors. Cytogenet Cell Genet. 51:985.

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. (1991) Nucleic Acids Research 19: 4008.

Donis-Keller et al(1987). A genetic linkage map of the human genome. Cell 51:319-337.

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

Edwards A, Civitello A, Hammond H A, and Caskey C T(1991a). DNA Typing and Genetic Mapping with Trimeric and Tetrameric Tandem Repeats. American Journal of Human Genetics 49:746-756.

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

Fidani L, and et al(1992). Screening for mutations in the open reading frame and promoter of the b-amyloid precursor protein gene in familial Alzheimer's disease: identification of a further family with APP717 Val -> Ile. Hum Molec Genet 1:

Freund AM, Bichara M, and Fuchs RPP(1989). Z-DNA forming sequences are spontaneous deletion hot spots. Proc Nat Acad Sci USA 86:7465-69.

Gardiner K, Horisberger M, Kraus J, Tantravahi U, Korenberg J, Rao V, Reddy S, Patterson D. (1990). Analysis of human chromosome 21: correlation of physical and cytogenetic maps; gene and CpG island distributions. EMBO. 9:25-34.

Gardiner K, Aissani B, Bernardi, G. (1990). A compositional map of human chromosome 21. EMBO. 9:1853-1858.

Goate A, Owen MJ, James LA, Mullan MJ, Rossor MN, Haynes AR, Farrall M, Lai LYC, oques P, Williamson R, Hardy JA. (1989). Predisposing locus for Alzheimer's disease on chromosome 21. Lancet 1 (8634):352-355.

Goate A, et al. & Hardy J. (1991) Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature 349:704-6.

Guo Z, Sharma V, Patterson D, and Litt M(1990). Dinucleotide Repeat Polymorphism At the D21S168 Locus. Nucleic Acids Research 18:5924.

Guo Z, Sharma V, and Litt M(1990). Dinucleotide Repeat Polymorphism At the D21S13E Locus. Nucleic Acids Research 18:7470.

Guo Z, Sharma V, Patterson D, Litt M. (1990) A Dinucleotide Repeat at the D21S167 Locus. Nucleic Acids Research 18:4967.

Hagelberg E, Gray I C, and Jeffreys A J(1991). Identification of the Skeletal Remains of a Murder Victim by DNA Analysis. Nature 352:427-429.

Haldane JBS. (1916) The combination of linkage values and the calculation of distances between the loci of linked factors. J. Genet. 8:299-309.

Hamada H, Kakunaga T (1982) Potential Z-DNA forming sequences are highly dispersed in the human genome. Nature 298: 396-8.

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.

Holliday R. (1989) Untwisting B-Z DNA. Trends in Genetics 5:355-6.

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

Hulten M, Lawrie N M, and Laurie D A(1990). Chiasma-Based Genetic Maps of Chromosome-21. American Journal of Medical Genetics 148-154.

Ichikawa H, Shimizu K, Saito A, Wang D, Oliva R, Kobayashi H, Kaneko Y, Miyoshi H, Smith C L, Cantor C R, and Ohki M(1992). Long-Distance Restriction Mapping of the Proximal Long Arm of Human Chromosome-21 with Not-I Linking Clones. Proceedings of the National Academy of Sciences of the United States of America 89:23-27.

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

Jefferys AL, Wilson V, Thein SL. (1985b) Individual-specific "fingerprints" of human DAN. Nature 316: 76-79.

Jefferys AL, et al. (1988) Spontaneous mutation rates to new length alleles at tandem repetitive hypervariable loci. Nature 332: 278-281.

Kamino K, Orr HT, Payami H, Wijsman EM, Anderson AD, O'dahl S, Nemens E, White JA, Sadovnick AD, Ball MJ, Warren A, Sharma V, Kukull W, Larson E, Heston LL, Martin GM, Bird TD, Schellenberg GD. (1992) Linkage and Mutational Analysis of Familial Alzheimer's Disiease Kindreds for the APP Gene Region. American Journal of Human Genetics, in press

Kao F T(1990). Microdissection and Microcloning of Human Chromosome-21. Molecular Genetics of Chromosome 21 and Down Syndrome 360:89-104.

Kazazian HH, Orkin SH, Markham AF, Chapman CR, Yousssoufian H, Waber PC. (1984) Quantification of close association between DNA haplotypes and specific ßthalassemia mutations in Mediterraneans. Nature 310:152-154.

Kidd KK, Bowcock AM, Schmidtke J, Track RK, Ricciuti F, Hutchings G, Bale A, Pearson P, Willard HF. (1989). Report of the committee and catalog of cloned and mapped genes and DNA polymorphisms. Ctytogenet. Cell Genet. 51:622-947.

Korenberg J R, Kawashima H, Pulst S M, Allen L, Magenis E, and Epstein C J(1990). Down Syndrome - Toward a Molecular Definition of the Phenotype. American Journal of Medical Genetics 91-97.

Kosambi DD. (1944) The estimation of map distances from recombination values. Ann. Eugen. 12:172-175.

Kwiatkowski D J, Henske E P, Weimer K, Ozelius L, Gusella J F, and Haines J(1992). Construction of a GT Polymorphism Map of Human 9q. Genomics 12:229-240.

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

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

Levinson G, Gutman GA. (1987) Slipped-strand mispairing: A major mechanism for DNA sequence evolution. Mol. Biol. Evol. 4(3): 203-221.

Lichter P, Ledbetter SA, Ledbetter DH, Ward WC. (1990). Fluorescence in situ hybridization with Alu and L1 polymerase chain reaction probes for rapid characterization of human chromosomes in hybrid cell lines. PNAS. 87:6634-6638.

Litt M. (1987). Notes on analysis of genetic linkage. (unpublished.)

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.

Litt M, Guo Z, Luo XY, Luty JA, Sharma V. (1990). Chromosomal distribution of highly polymorphic TG microsatellites. Am. J. Hum. Genet. Suppl. 47:A189.

Litt M. (1991a). "PCR of microsatellites." In: Polymerase Chain Reaction- A Practicle Approach; IRL Press. Oxford University Press.

Litt M, Sharma V, Luty JA. (1989) A Highly Polymorphic (TG)n Microsatellite at the D11S35 Locus. Cytogenet. Cell Genet. 51:1034

Ludecke HJ, Senger G, Claussen U, Horshemke B. (1989). Cloning defined regions of the human genome by microdissection of banded chromosomes and enzymatic amplification. Nature 338:348-350.

Luty JA, Guo Z, Willard HF, Ledbetter DH, Ledbetter S, Litt M. (1990). Five polymorphic microsatellite VNTRs on the human X chromosome. Am. J. Hum. Genet. 46:776-783.

Maniatis T, Fritsch E, Sambrook J. (1982) Molecular Cloning: A Laboratory Handbook. Cold Spring Harbor, New York.

Maslen CL. (1987) The cloning and characterization of DNA probes which detect restriction length polymorphisms on human chromosome 11q qnd 16q. Ph.D. dissertation, Oregon Health Sciences University.

McCormic MK et al & Antonarakis (1989) Construction of human chromosome 21 specific yeast artificial chromosomes. Proc. Natl. Acad. Sci.USA 86: 9991-5.

McInnis MG, Blaschak J, Cox T, Avramopoulos D, Bluin JL, Kalaitsidaki M, Sharma V, Petersen MB, Warren AC, van Broeckhoven C, Litt M, Chakravarti A, Antonorakis SE. (1992) Linkage map of chromosome 21 using DNA polymorphic markers genotyped by PCR. Cytogenet. Cell. Genet., in press

McKusick VA. (1990) Mendelian Inheritance of Man. 9th ed. Baltimore: Johns Hopkins University Press.

Meyers RM, Larin Z, Maniatis T. (1985a) Detection of single base substitutions by RNAse cleavage at mismatches in RNA:DNA hybrids. Science 230:1244-6.

Meyers RM , Maniatis T. (1986) Recent advances in the development of methods for detecting single base substitutions associated with human genetic diseases. Cold Spring Harbor Symposium on Quantitative Biology LI: 275-284.

Micklos DA, Freyer GA. (1990) DNA Science, A First Course in Recombinant DNA Technology, Carolina Biologocal Supply Company, Burlington, NC

Mohandas T. *et al.*, (1986) Somatic Cell. Molec. Genet. 12: 89-94.

Morral N, Nunes V, Casals T, and Estivill X(1991). CA/GT Microsatellite Alleles Within the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Gene Are Not Generated by Unequal Crossingover. Genomics 10:692-698.

Moore S S, Sargeant L L, King T J, Mattick J S, Georges M, and Hetzel D J S (1991). The Conservation of Dinucleotide Microsatellites Among Mammalian Genomes Allows the Use of Heterologous PCR Primer Pairs in Closely Related Species. Genomics 10:654-660.

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

Neve RL, Stewart GD, Newcomb P, Van Keuren ML, Patterson D, Drabkin HA, Kurnit DM. (1986). Human chromosome 21 cDNA clones. Gene. 49:361-369.

Nizetic D, Zehetner G, Monaco A P, Gellen L, Young B D, and Lehrach H(1991). Construction, Arraying, and High-Density Screening of Large Insert Libraries of Human Chromosome-X and Chromosome-21 - Their Potential Use As Reference Libraries. Proceedings of the National Academy of Sciences of the United States of America 88:3233-3237.

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single strand conformation polymorphisms. Proc. Natl. Acad. Sci.USA 86: 2766-2770.

Ostrander E A, Jong P M, Rine J, and Duyk G(1992). Construction of Small-Insert Genomic DNA Libraries Highly Enriched for Microsatellite Repeat Sequences. Proc. Natl. Acad. Sci. USA 89:3419-3423.

Ott J (1991) Analysis of Human Genetic Linkage. Rev. ed. Johns Hopkins University Press.

Owen MJ, James LA, Hardy JA, Williamson R, Goate AM. (1990). Physical mapping around the Alzheimer's disease locus on the proximal long arm of chromosome 21. Am. J. Hum. Genet. 46:316-322.

Patterson D, Schild D, and Lai L W(1990). Translocations and Rearrangements Involving Chromosome-21. Molecular Genetics of Chromosome 21 and Down Syndrome 360:27-37.

Payami H. (1984). " Disease studies: the classical affected sib-pair method." Doctoral Dissertation, UC Berkeley. pgs 8-15

Payami H. (1990). HLA and genetic susceptibility to Alzheimer's disease. Dept. of Health Human Services/Public Health Service Project Proposal.

Pericak-Vance MA, Bebout JL, Yamoaka LA, Gaskell PC Jr., Hung W-Y, Alberts MJ, Clark CM, Haynes CS, Welsh KA, Earl NL, Heyman A, Roses AD. (1990). "Linkage studies in familial Alzheimer's disease , application of the affected pedigree member method of linkage analysis." pgs. 215-228, In: Molecular Biology and Genetics of Alzheimer's Disease. Elsevier Science Publishers B.V. (Biomedical Division).

Petersen M B, Schinzel A A, Binkert F, Tranebjaerg L, Mikkelsen M, Collins F A, Economou E P, and Antonarakis S E(1991). Use of Short Sequence Repeat DNA Polymorphisms After PCR Amplification to Detect the Parental Origin of the Additional Chromosome-21 in Down Syndrome. American Journal of Human Genetics 48:65-71.

Petersen M B, Slaugenhaupt S A, Lewis J G, Warren A C, Chakravarti A, and Antonarakis S E (1991). A Genetic Linkage Map of 27 Markers on Human Chromosome-21. Genomics 9:407-419.

Saiki RK et al., & Erlick R. (1988) Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487-491.

Schellenberg GD, Bird TD, Wijsman EM, Moore DK, Boehnke M, Bryant EM, Lampe TH, Nochlin D, Sumi SM, Deeb SS, Beyreuther K, Martin GM. (1988), Absence of linkage of chromosome 21q21 markers to familial Alzheimer's disease. Science. 241:1507-1510.

Schellenberg G D, Pericakvance M A, Wijsman E M, Moore D K, Gaskell P C, Yamaoka L A, Bebout J L, Anderson L, Welsh K A, Clark C M, Martin G M, Roses A D, and Bird T D(1991). Linkage Analysis of Familial Alzheimer Disease, Using Chromosome-21 Markers. American Journal of Human Genetics 48:563-583.

Sharma V, & Litt M. (1992) A Tetrameric Repeat at D21S11. Human Molecular Genetics Vol 1:69.

Sharma V, de Jong P, Litt M. (1992) A Dinucleotide Repeat at D21S236. Human Molecular Genetics, Vol 1, No. 4, pg. 289.

Sharma V, Guo Z, Litt M. (1992) A Dinucleotide Repeat at D18S37. Human Molecular Genetics, Vol 1, No. 4, pg. 289.

Sharma V, de Jong P, Litt M. (1992) A Dinucleotide Repeat at D21S218. (submitted)

Sharma V, & Litt M. (1992) A Dinucleotide Repeat at D4S233. Human Molecular Genetics, (submitted)

Sharma V, Allen L, Magenis RE, Litt M. (1991) A Dinucleotide Repeat at the D21S172 Locus. Nucleic Acids Research 19:1169

Sharma V, Allen L, Magenis RE, Litt M. (1991) A Dinucleotide Repeat at the D1S116 Locus. Nucleic Acids Research 19:1169

Sharma V, Litt M. (1991) A Dinucleotide Repeat at D1S117. Nucleic Acids Research 19:1168

Sharma V, Smith L, Allen L, Magenis RE, Litt M. (1991) A Dinucleotide Repeat at the D14S43 Locus. Nucleic Acids Research 19:1722

Sharma V, Smith L, Magenis RE, Litt M. (1991) A Dinucleotide Repeat at the D9S55 Locus. Nucleic Acids Research 19:4023

Sharma V, Luty JA, Kramer P, Billingsley G, Cox D, Litt M. (1991) Addition of Two New Highly Informative (CA)n Repeat Polymorphisms to the Index Map of Chromosome 14. Abstract and Poster presented at Human Genome Mapping Conference 11, CytogenLondon UK

Siddique T, and et al(1991). Linkage of a gene causing familial amyotrophic lateral sclerosis to chromosome 21 and evidence of genetic-locus heterogeneity. New England J Med 324:1381-1384.

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 CAB. (1986) The development of human linkage analysis. Ann. Hum. Genet. 50:293-311.

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.

Stephens J C, Cavanaugh M L, Gradie M I, Mador M L, and Kidd K K(1990). Mapping the Human Genome - Current Status. Science 250:237-244.

Stewart G D, and Kurnit D M(1990). Recombination-Based Screening for Genes on Chromosome-21. American Journal of Medical Genetics 115-119.

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

St. George-Hyslop PJ, Tanzi RE, Polinsky RJ, Neve RL, Pollen D, Drachman D, Growdon J, Cupples LA, Nee L, Myers RH, O'Sullivan D, Watkins PC, Amos JE, Deutsch CK, Bodfish JW, Kinsbourne M, Feldman RG, Bruni A, Amaducci L, Foncin JF, Gusella JF. (1987). Absence of duplication of chromosome 21 genes in familial and sporadic Alzheimers disease. Science 238:664-666.

St. George-Hyslop PH, et al (1990). Genetic linkage studies suggest that Alzheimer's disease is not a single homogeneous disorder. Nature. 347:194-197.

Stinissen P, and Van Broeckhoven C(1991). A New (CA)n Repeat Polymorphism at the D21S13E Locus. Nucleic Acids Research 19:5089.

Tantravahi U, Stewart GD, Van Keuren M, McNeil G, Roy S, Patterson D, Drabkin H, Lalande M, Kurnit DM, Latt SA. (1988). Isolation of DNA sequences on human chromosome 21 by application of a recombination-based assay to DNA from flow sorted chromosomes. Hum. Genet. 79:192-202.

Tanzi R E, Vaula G, Romano D M, Mortilla M, Huang T L, Tupler R G, Wasco W, Hyman B T, Haines J L, Jenkins B J, Kalaitsidaki M, Warren A C, Mcinnis M C, Antonarakis S E, Karlinsky H, Percy M E, Connor L, Growdon J, Crappermclachlan D R, Gusella J F, and Stgeorgehyslop P H(1992). Assessment of Amyloid beta-Protein Precursor Gene Mutations in a Large Set of Familial and Sporadic Alzheimer Disease Cases. American Journal of Human Genetics 51:273-282.

Tanzi RE, St. George-Hyslop PH, Haines JL, Polinsky RJ, Nee L, Foncin JF, Neve RL, McClatchey AI, Conneally PM, Gusella JF. (1987). The genetic defect in familial Alzheimer's disease is not tightly linked to the beta-amyloid gene. Nature 329:156-157.

Tanzi R E, Haines J L, and Gusella J F(1990). Detailed Genetic Linkage Map of Human Chromosome-21 - Patterns of Recombination According to Age and Sex. Molecular Genetics of Chromosome 21 and Down Syndrome 360:15-26.

Tanzi RE, St George-Hyslop P, Gusella JF. (1991a) Molecular Genetics of Alzheimer disease amyloid. J. Biol. Chem. 266: 20579-20582.

Tautz D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research 17: 6463-71.

Threadgill et al (1991): Evidence for the evolutionary origin of human chromosome 21 from comparative gene mapping in the cow and mouse. PNAS 88:154-158.

Stinissen P, and Van Broeckhoven C(1991). A New (CA)n Repeat Polymorphism at the D21S13E Locus. Nucleic Acids Research 19:5089.

Tantravahi U, Stewart GD, Van Keuren M, McNeil G, Roy S, Patterson D, Drabkin H, Lalande M, Kurnit DM, Latt SA. (1988). Isolation of DNA sequences on human chromosome 21 by application of a recombination-based assay to DNA from flow sorted chromosomes. Hum. Genet. 79:192-202.

Tanzi R E, Vaula G, Romano D M, Mortilla M, Huang T L, Tupler R G, Wasco W, Hyman B T, Haines J L, Jenkins B J, Kalaitsidaki M, Warren A C, Mcinnis M C, Antonarakis S E, Karlinsky H, Percy M E, Connor L, Growdon J, Crappermclachlan D R, Gusella J F, and Stgeorgehyslop P H(1992). Assessment of Amyloid beta-Protein Precursor Gene Mutations in a Large Set of Familial and Sporadic Alzheimer Disease Cases. American Journal of Human Genetics 51:273-282.

Tanzi RE, St. George-Hyslop PH, Haines JL, Polinsky RJ, Nee L, Foncin JF, Neve RL, McClatchey AI, Conneally PM, Gusella JF. (1987). The genetic defect in familial Alzheimer's disease is not tightly linked to the beta-amyloid gene. Nature 329:156-157.

Tanzi R E, Haines J L, and Gusella J F(1990). Detailed Genetic Linkage Map of Human Chromosome-21 - Patterns of Recombination According to Age and Sex. Molecular Genetics of Chromosome 21 and Down Syndrome 360:15-26.

Tanzi RE, St George-Hyslop P, Gusella JF. (1991a) Molecular Genetics of Alzheimer disease amyloid. J. Biol. Chem. 266: 20579-20582.

Tautz D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research 17: 6463-71.

Threadgill et al (1991): Evidence for the evolutionary origin of human chromosome 21 from comparative gene mapping in the cow and mouse. PNAS 88:154-158.

Warburton PE, Greig GM, Haaf T, Willard HF. (1991) PCR amplification of chromosome specific alpha satellite DNA: Definition of centromeric STS markers and polymorphic analysis. Genomics 11: 324-333.

Warren AC, Slaugenhaupt SA, Lewis JG, Chakravarti A, Antonarakis SE. (1989). A genetic linakge map of 17 markers on human chromosme 21. Genomics. 4:579-591.

Watkins P C(1990). Isolation and Characterization of DNA Probes for Human Chromosome-21. Molecular Genetics of Chromosome 21 and Down Syndrome 360:53-67.

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.

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

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

White R, Caskey CT. (1988) The human as an experimental system in molecular genetics. Science 240: 1483-1488.

Winterø KW, Fredholm M, Thomsen PD. (1992) Variable (dG-dT)n/(dC-dA)n sequences in the porcine genome. Genomics 12: 281-8.

Wolff R, Plaetke R, Jefferys AJ, White R. (1989) Unequal crossingover between homologous chromosomes is not the major mechanism involved in the generation of new alleles at VNTR loci. Genomics 5: 382-4.

Wood *et al.* (1986) Proc. Natl. Acad. Sci. USA 82: 1585-88

Wu DY, Ugozzoli L, Pal BK, Qian J, Wallace RB. (1991) The effect of temperature and oligonucleotide primer length on the specificity and efficiency of amplification by the polymerase chain reaction. DNA and Cell Biology 10: 233-238.

Yu J W, Hartz J, Xu Y S, Gemmill R M, Korenberg J R, Patterson D, and Kao F T(1992). Isolation, Characterization, and Regional Mapping of Microclones from a Human Chromosome-21 Microdissection Library. American Journal of Human Genetics 51:263-272.

VII: APPENDIX A

American Journal of Human Genetics (1990), Supplement 47:A189

Chromosomal distribution of highly polymorphic TG microsatellites.M. Litt, Z. Guo, X.Y. Luo, J. Luty and V. Sharma . Oregon Health Sciences University, Portland.

More than 3000 RFLPs have been described and linkage maps have been published for most human chromosomes. However, the small polymorphism information content (PIC) of most known marker loci is often a limiting factor in linkage analysis of inherited diseases. Although the discovery of minisatellites, also known as variable number of tandem repeat (VNTR) polymorphisms has helped to reduce this limitation, highly informative VNTR loci have not been found on all chromosome arms and those which have been identified are often localized near telomeres, leaving large regions of the genome out of reach of multiallelic marker loci.

TG microsatellites, also known as CA repeats, are a class of highly repeated interspersed sequences in the human and other eukaryotic genomes. They often display length polymorphism of the VNTR type and they can be typed with the polymerase chain reaction (PCR) using unique sequences that flank the repeats as primers and resolving the products on DNA sequencing gels (Weber and May, AJHG 44:388, 1989; Litt and Luty, AJHG 44:397,1989). An important question which bears on the usefulness of TG microsatellites is their chromosomal distribution. Do they, like minisatellite-based VNTRs, tend to localize near telomeres, or is their distribution more nearly uniform? To elucidate this question, we have characterized 16 TG microsatellites with respect to both PIC and chromosomal localization. Thirteen of these loci were polymorphic, with PICs ranging between 0.20 and 0.88 (average PIC=0.67). Chromosomal localizations of 11 of these VNTRs clearly exclude them from telomeric G-bands, whereas the remaining two have only been assigned to whole chromosomes. Of 14 mapped TG microsatellite VNTRs characterized by other groups, 8 can clearly be excluded from telomeric G-bands and none have as yet been unambiguously assigned to such regions. Hence, we conclude that, in contrast to minisatellite VNTRs, the chromosomal distribution of TG microsatellite VNTRs is not biased towards telomeres. Therefore, this class of markers should be very useful for filling gaps in human linkage maps as well as for development of highly informative loci closely flanking disease genes. Because of their wide genomic distribution and high PICs, TG microsatellite VNTRs will also be very useful for affected sib-pair and affected relative pair analyses of inherited human diseases.

VII: APPENDIX B

American Journal of Human Genetics (1992), Supplement in press

Linkage map of human chromosome 21 using DNA polymorphic markers genotyped by PCR. MG McInnis[1], J Blaschak[2], T Cox[2], D Avramopoulos[1], JL Blouin[1], M Kalaitsidaki[1], V Sharma[3], MB Petersen[4], AC Warren[1], C van Broeckhoven[5], M Litt[3], A Chakravarti[2], SE Antonarakis[1]. [1]Center for Medical Genetics, Johns Hopkins University School of Medicine, Baltimore MD, [2]University of Pittsburgh, Pittsburgh, PA, [3]Oregon Health Sciences University, Portland, OR, [4]The John F Kennedy Institute, Glostrup Denmark, [5]University of Antwerp (UIA), Belgium.

We have created a genetic linkage map for human chromosome 21 (HC 21) using 40 CEPH pedigrees and DNA polymorphic markers genotyped by PCR. This map contains 38 polymorphic loci, consisting of (GT)n, (TAAA)n and (TCTA)n repeats or polymorphisms found in the 3' untranslated regions of genes. Twenty four loci have heterozygosity >60%. Nine genes are included in the map (GART, AML1, PFKL, CBS, CBR, S100b, IFNAR, CD18 & HMG14). After the initial CEPH genotyping, the data were examined with the computer program "Chrompic" and double recombinants due to potential data errors were reanalyzed. Twenty six markers were placed in the map with odds > 1000:1. The total sex averaged length of the long arm of HC 21 (from D21S215 to D21S171) is 77.9 cM assuming Kosambi interference. The male map is 59.5 cM while the female map is 97.0 cM. The male/female difference is due to an increased recombination in females at the proximal short arm interval between D21S215 and D21S214. The average distance between adjacent loci is 2 - 3 cM. This index linkage map of HC 21 provides a basis for linkage studies of disease phenotypes for this chromosome.

VII: APPENDIX C

Science (1992) in press

Chromosome 21

S. E. Antonarakis[a], A. Chakravarti[b], M.G. McInnis[a], V. Sharma[c], D. Avramopoulos[a], J. E. Blaschak[b], M. Litt[c], J-L. Blouin[a]

The human chromosome 21 map represents a subset of 13 evenly spaced and highly informative polymorphic loci selected from a collection of more than 50 genetic markers. A comprehensive map of 42 polymorphic markers genotyped in CEPH pedigrees by PCR will be reported elsewhere (*87*). The map consists of 13 loci, of which 11 can be genotyped by PCR-based methods. The heterozygosity of these markers ranges from 60 to 95%, with 12 of the 13 markers having heterozygosity greater than 70%. The chromosome 21 map has a male genetic length of 57 cM, a female genetic length of 78 cM and a sex-average genetic length of 67 cM. On the basis of cytogenetic localization of the terminal loci (D21S120 and D21S112), we estimate that the genetic map spans almost the entire physical length of the long arm of the chromosome (39 Mb) (*38*). On the sex-average map the 12 marker intervals range from 3.2 cM to 7.9 cM, with an average spacing of 5.6 cM. Sex-specific differences in recombination frequency were examined along the chromosome arm by testing each interval. Significant sex-differences were observed at the terminal intervals: D21S120-D21S11 where females showed greater recombination than males, and, for the interval D21S171-D21S112 where males showed greater recombination than females. There appears to be excessive recombination towards the telomere of the chromosome. Linkage maps for chromosome 21 that contain polymorphic markers typed by Southern blotting in the CEPH or the Venezuelan reference pedigrees have been reported (*88*). The map presented here is considerably shorter than the previously published maps in CEPH pedigrees, presumably because rigorous error checking has eliminated false recombination events. The current map includes approximately 95% of the physical length of the chromosome and is comprised of highly informative, closely spaced markers.

VII: APPENDIX D

Cytogenet. Cell. Genet. (1991) 58: 1989

ADDITION OF TWO NEW HIGHLY INFORMATIVE (CA)n REPEAT POLYMORPHISMS TO THE INDEX MAP OF CHROMOSOME 14. V. SHARMA, J.A. LUTY, P. KRAMER, G. BILLINGSLEY*, D. COX* and M. LITT. Department of Medical Genetics, Oregon Health Sciences University, Portland, OR 97201 and * Hospital for Sick Children, Toronto, ONT M5G 1X8, Canada.

A short-term goal of the US Human Genome Initiative is the construction of index maps for all human chromosomes, consisting of markers with minimum heterozygosities of 70% spaced at intervals not greater than 15 cM. Recently, we described D14S43 and D14S45, two highly informative (CA)n repeat markers on chromosome 14 (1,2). We have now typed these markers in CEPH families thereby adding these markers to the index map of this chromosome.

D14S43 has 10 alleles and a PIC of 0.72. It has been mapped to 14q24.3 by flourescent in situ hybridization (1). We have typed this marker in 37 informative CEPH families. Pairwise linkage studies with FOS in all informative CEPH families gave a maximum LOD score of 23.8 at theta=0. D14S43 was also linked to SPTB: pairwise analysis in 21 informative CEPH families gave a maximum LOD score of 17.7 at theta=0.15.

D14S45 has 9 alleles and a PIC of 0.74 (2). Three-point linkage analysis in 23 CEPH families gives a maximum liklihood map: 14cen-PI-16cM-D14S45-16cM-D14S13-14qter. Odds against inversion of adjacent loci were greater than 10,000:1.

Both of these new microsatellite markers fill gaps in the index map of chromosome 14 and will therefore be very useful for identification og disease genes which may be present on this chromosome.

(Supported in part by NIH Grant HG00022. V. Sharma is the recipient of a fellowship from the Medical Research Foundation of Oregon.)

HGM II, London 1991

Poster presented at HGM 11 (1991)

Introduction: A short term goal of the US Human Genome Initiative is the construction of index maps for all human chromosomes, consisting of markers with minimum heterozygosities of 70% spaced at intervals not greater than 15 cM. Recently, we described D14S43 and D14S45, two highly informative (CA)n repeat markers on chromosome 14 (1,4). We have now typed these markers in CEPH families thereby adding them to the genetic linkage map of this chromosome.

D14S43

Source and Description: Cosmid ICRFc102E1222 was from a flow sorted chromosome 21 cosmid library. DNA sequences flanking the sequence $(TG)_{14}$ TTGGGAATGGGGCTGGGGGGAGACAGGGT$(TG)_{10}$ contained in HaeIII subclone 2E12B of this cosmid were used to design PCR primers.

PCR Primers: 2E12B#1: 5'-TGGAACACTCAGGCGA-3'
2E12B#2: 5'-CCAGAGCCACTTTCTAC-3'

Polymorphism: Ten allelic fragments were resolved on DNA sequencing gels. Lengths of allelic fragments (nt) were: A1=193, A2=191, A3=189, A4=187, A5=185, A6=183, A7=181, A8=161, A9=159, A10=155.

Frequencies: Allele frequencies in 52 unrelated individuals were: A1= .01, A2= .01, A3= .11, A4= .13, A5= .26, A6= .02, A7= .06, A8= .01, A9= .38, A10= .01; PIC=0.72

Chromosome Location: D14S43 has been mapped to 14q24.3 by fluorescent in situ hybridization (1).

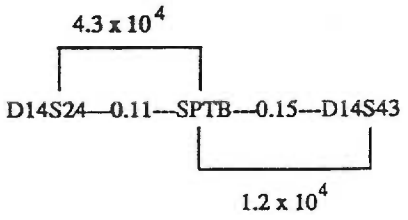
PCR Conditions: PCR reactions are carried out in a total volume of 25 ul containing: 25 ng genomic DNA, 10 pmole of each primer, 1.5 mM $MgCl_2$, 200 µM dNTPs, 50 mM KCl, 10 mM Tris-Cl, pH 8.3, 0.6 units Taq polymerase (Ampli-Taq) and 0.01% gelatin. Amplification is for 35 cycles with denaturation at 94 C for 60 seconds, annealing at 54 C for 60 seconds and extension at 72 C for 30 seconds. PCR products are resolved on DNA sequencing gels, capillary-blotted onto Hybond N+ nylon membranes and revealed by probing with kinased (CA)15 oligomer.

Linkage analyses: We typed 37 informative CEPH families at D14S43. Mendelian inheritance was observed in all cases. Using data from the CEPH database v4 for D14S24 (CRI-C70) , PI and AACT as well as our own data for FOS and SPTB, we calculated the following two point LOD scores for D14S43 versus the indicated markers:

Theta	0.00	0.05	0.10	0.20	0.30	$\hat{\theta}$	$\hat{Z}$
FOS	23.8	22.7	19.5	14.7	9.5	0.00	23.8
SPTB	-∞	12.1	16.8	16.9	12.7	0.15	17.7
D14S24	-∞	-2.9	3.5	7.1	6.6	0.23	8.3
PI	-∞	-18.9	-6.9	2.0	4.1	0.31	4.1
AACT	-∞	-12.2	-4.4	1.1	2.3	0.30	2.3

Probe pALP13-8 at the FOS locus detects TaqI & BstNI RFLPs which when haplotyped give a combined PIC of 0.243. D14S43 is tightly linked to FOS at $\hat{\theta}$=0. Since FOS has been previously localized to q24.3 (2), this confirms localization of D14S43 to the same region. Probe pB8 at the ß-spectrin locus (SPTB) detects a three allele, HindIII RFLP with a PIC of 0.36 (3,5) and is also linked to D14S43. Significant linkage is also found with D14S24 (CRI-C70/MspI) with a maximum LOD score of 8.3 at θ= 0.23.

Three-point linkage analysis in these CEPH families gave the following maximum likelihood map. Distances between loci are shown as recombination fractions; odds against inversion of adjacent loci are also shown and exceed 10,000:1.



VII: APPENDIX E

American Journal of Human Genetics (1992) in press

Linkage and Mutational Analysis of Familial Alzheimer Disease Kindreds for the APP Gene Region.

Running Title: The APP Gene and Familial Alzheimer Disease

Authors: Kouzin Kamino[1, 2, 3], Harry T. Orr[4], Haydeh Payami[5], Ellen M. Wijsman[6], Stefan M. Pulst[7], Leojean Anderson[1], Sheldon O'dahl[1], Ellen Nemens[1], June A. White[8], Adele D. Sadovnick[9], Melvyn J. Ball[10], Jeffery Kaye[11], Andrew Warren[12, 13], Melvin McInnis[12, 13], Stylianos E. Antonarakis[12], Julie R. Kornbreg[7], Vikram Sharma[5], Walter Kukull[14], Eric Larson[15], Leonard L. Heston[16], George M. Martin[2], Thomas D. Bird[1, 17] and Gerard D. Schellenberg[1]

[1] Division of Neurology, Department of Medicine, University of Washington, Seattle

[2] Department of Pathology, University of Washington, Seattle, Washington

[3] Present address; Department of Medicine and Geriatrics, Osaka University Medical School, Osaka, Japan

[4] Department of Laboratory Medicine and Pathology and the Institute of Human Genetics, University of Minnesota, Minneapolis, Minnesota

[5] Department of Molecular and Medical Genetics, Oregon Health Sciences University, Portland, Oregon

[6] Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, Washington

[7] Division of Neurology, Cedars-Sinai Medical Center, University of California, Los

Angeles

[8] Department of Psychiatry, University of Minnesota, Minneapolis, Minnesota

[9] Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

[10] Division of Neuropathology, Oregon Health Sciences University, Portland, Oregon

[11] Department of Neurology, Oregon Health Sciences University, Portland, Oregon

[12] Center for Medical Genetics, Johns Hopkins University of Medicine, Baltimore, MD

[13] Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, MD

[14] Department of Epidemiology, University of Washington, Seattle, Washington

[15] Department of Medicine, University of Washington, Seattle

[16] The Washington Institute for Mental Illness Research and Training, Department of Psychiatry, University of Washington

[17] Division of Neurology, Veterans Administration Medical Center, Seattle

Correspondence and Reprint Requests: Gerard D. Schellenberg, Division of Neurology RG-27, University of Washington, Seattle, WA, 98195, (206) 543-2340

SUMMARY

A large number of familial Alzheimer disease (AD) kindreds were examined to determine whether mutations in the amyloid precursor protein (APP) gene could be responsible for the disease. Previous studies have identified 3 mutations at APP codon 717 which are pathogenic for AD. Samples from affected subjects were examined for mutations in exons 16 and 17 of the APP gene. A combination of direct sequencing and single strand conformational polymorphism (SSCP) analysis was used. Sporadic AD and normal controls were also examined by the same methods. Five sequence variants were identified. One variant at APP codon 693 resulted in a Glu→Gly change. This is the same codon as the Hereditary Cerebral Hemorrhage with Amyloidosis-Dutch type Glu→Gln mutation. Another single base change at APP codon 708 did not alter the amino acid encoded at this site. Two point mutations and a 6 bp deletion were identified in the intronic sequences surrounding exon 17. None of the variants could be unambiguously determined to be responsible for FAD. The larger families were also analyzed by testing for linkage of FAD to a highly polymorphic short tandem repeat marker (D21S210) which is tightly linked to APP. Highly negative LOD scores were obtained for the family groups tested and linkage was formally excluded beyond $\Theta = 0.10$ for the Volga German kindreds, $\Theta = 0.20$ for early-onset non-Volga Germans, and $\Theta = 0.10$ for late-onset families. Lod scores for linkage of FAD to markers centromeric to APP (D21S1/S11, D21S13, and D21S215) were also negative in the 3 family groups. These studies show that APP mutations account for AD in only a small fraction of FAD kindreds.

INTRODUCTION

The amyloid precursor protein (APP) gene encodes for at least 4 isoform proteins (Kang et al. 1987; Tanzi et al. 1988; Ponte et al. 1988; Kitaguchi et al. 1988; DeSauvage and Octave, 1989) which are found in many tissues including the central nervous system (Tanzi et al. 1987). Although the normal functions of these proteins are not presently understood, at least 2 isoforms contain functional protease inhibitor domains (Tanzi et al. 1988; Ponte et al. 1988; Kitaguchi et al. 1988). In Alzheimer disease (AD), one or more forms of APP is cleaved to yield a 39-43 amino acid fragment (Glenner and Wong, 1984) called the β-amyloid peptide. This fragment is found in amyloid deposits associated with the cerebral vascular system and with neuritic plaques.

Recently Goate et al. (1991) identified a mutation in the APP gene at codon 717 which results in a Val→Ile substitution in the region adjacent to the C-terminal end of the β-amyloid peptide. The mutation co-segregated with familial Alzheimer disease (FAD) in 2 families, providing strong evidence that the mutation is pathogenic for AD. This conclusion was supported by identification of 4 other families in which the same mutation co-segregated with the disease (Naruse et al. 1991; Yoshioka et al. 1991). The codon 717 mutation was not found in a large number of sporadic AD (spAD) subjects or normal controls (Goate et al. 1991; Van Duijn et al. 1991; Naruse et al. 1991; Yoshioka et al. 1991; Schellenberg et al. 1991a; Chartier-Harlin et al. 1991a) . Two additional mutations in the same codon were found in 2 other FAD kindreds by Murrell

et al. (1991) and Chartier-Harlin et al. (1991b) which result in Val→Phe and Val→Gly amino acid substitutions, respectively. Thus, in at least some families, a mutation at codon 717 is sufficient to cause early-onset FAD.

Numerous early and late-onset families have been screened for the 717 Ile variant (Goate et al. 1991; Van Duijn et al. 1991; Naruse et al. 1991; Yoshioka et al. 1991; Schellenberg et al. 1991a; Chartier-Harlin et al. 1991a) and only 6 kindreds have this mutation. In the following work, a large number of both early and late-onset kindreds were evaluated to determine whether codon 717 and other mutations in the APP gene could be responsible for FAD in these families. Linkage analysis of FAD to a highly polymorphic marker D21S210 closely linked to APP was evaluated to attempt to identify what fraction of FAD families can be expected to have APP mutations at locations other than codon 717. Since many kindreds are too small to be evaluated by linkage methods, we also used direct sequence analysis and single strand conformational polymorphism (SSCP) analysis to screen for known mutations and to look for new variants in exons 16 and 17 of the APP gene (the region encoding the β-amyloid peptide).

MATERIALS AND METHODS

FAD Families. Families were evaluated either by the University of Washington Alzheimer's Disease Research Center, the Oregon Health Sciences University Alzheimer's Disease Center Core or the University of Minnesota Alzheimer's disease

research group as part of a consortium for the study of familial Alzheimer's disease. Affected subjects were evaluated and diagnosis of AD assigned as previously described (Schellenberg et al. 1991b; McKhann et al. 1984). Autopsies were performed as previously described (Bird et al. 1988; Bird et al. 1989). The following families (with observed family mean age-of-onset given in parenthesis) were used for linkage analysis and have been previously described; AM (42 $\pm$ 2.5), BMH (76 $\pm$ 7), CK (65 $\pm$ 6.8), CSF (70 $\pm$ 5.3), E (57 $\pm$ 4.5), H (58 $\pm$ 1.0), HB (60 $\pm$ 7.2), HD (59 $\pm$ 10), JR (78 $\pm$ 1.9), KG (44 $\pm$ 1.0), KS (64 $\pm$ 5.2), L (41 $\pm$ 4.6), MI (70 $\pm$ 4.0), MMM (64 $\pm$ 0.5), P (67 $\pm$ 4.4), R (51 $\pm$ 7.1), SNW (52 $\pm$ 2.5), T (66 $\pm$ 4.8), V (46 $\pm$ 3.8), W (54 $\pm$ 3.9), WLA (68 $\pm$ 4.0), 603 (48 $\pm$ 6.5) (Bird et al. 1988; Bird et al. 1989; Cook et al. 1979; Goudsmit et al. 1981; Schellenberg et al. 1991a; Schellenberg et al. 1991b), HR-A9 (64 $\pm$ 17), HR-I (47 $\pm$ 4.6), HR-X (58 $\pm$ 2.3), HR-XII (68 $\pm$ 6.1), HR-XIII (47 $\pm$ 3.6) (Heston et al. 1991), and HR-XV (42 $\pm$ 3.9)(Valencia et al. 1986). The characteristics of the families not previously described are shown in Table 1. Families used only for mutational analysis are listed in Table 1 and in Table 1 of Schellenberg et al. (1991a). In addition to FAD kindreds, affected subjects from 5 families with mixed familial neurodegenerative disorders potentially related to AD (e.g. LIT family) were used in the mutational analysis.

Sporadic AD and Control Subjects. The spAD group consisted of 153 consecutive newly diagnosed AD subjects meeting the NINCDS criteria for probable AD (mean age-of-onset M = 75.6, SD = 6.62) ascertained through a Seattle area health care maintenance organization (Group Health Cooperative of Puget Sound). A randomly chosen control

group from the same population consisted of 207 subjects (mean age = 78.0, SD = 6.08). These subjects had Mini-Mental State Exam scores $\geq$ 28 ($\geq$ 27 if over 80 yrs), no indication of dementia on the Mattis Dementia Rating Scale or the Blessed Dementia Rating scale, and no history of dementia determined by interview or medical record review. A subset of this population was described previously (Schellenberg et al. 1991a). A second control group consisted of 100 subjects who responded to an advertisement for blood donors (Schellenberg et al. 1991a).

Genotype Analysis. Southern blot analysis was performed as previously described (Schellenberg et al. 1991b). The clone for D21S16 (pGSE9) was obtained from G. Stewart, a partial cDNA clone for the APP gene from K. Beyreuther, D21S13 (G21RK, GSM21) from the ATCC clone bank, D21S1/11 (formerly D21S72) (pG95-alpha1-11a, pG95-alpha1-11b, and pG95-alpha1-11c) from B.N. White, and D21S1 (pPW228C) from Paul Watkins. RFLP polymorphisms typed were TaqI, PstI, EcoRI, and HaeIII for D21S13, MspI, BamHI, TaqI, HincI, and PvuII for D21S1/S11, and EcoRI and BglII for APP as described in by Williamson et al. (1991). Di and tetra nucleotide repeat polymorphisms were genotyped as previously described (Schellenberg et al. 1992). The primers and amplification conditions used are as described by others for D21S210 (Warren et al. 1992a), D21S215 (Warren et al. 1992b), D21S13 (Guo et al. 1990; Stinissen and Van Broeckhoven 1991), and D21S11 (Sharma and Litt 1992).

Linkage Analysis. LOD scores were calculated as previously described (Schellenberg et

al. 1991b; 1992) using the computer program LIPED modified to handle up to 10 alleles. A cumulative normal age-of-onset correction was used (Schellenberg et al. 1991b). Autosomal dominant inheritance with age-dependent penetrance was assumed. The age curve was constructed as a cumulative normal distribution using family-specific means, and an overall standard deviation of 10.87 or 11.99 years. LOD scores were also calculated by setting the penetrance of the AD genotype to a fixed penetrance of 1% which effectively results in ignoring the AD phenotypes of the unaffected but at-risk subjects. Multipoint analysis was performed as previously described (Schellenberg et al. 1991b) using the LINKMAP program of the LINKAGE package (Lathrop et al. 1984). Possible heterogeneity among groups of families was examined using the μ-test (Morton, 1956). Groups compared were Volga German versus early-onset non-Volga Germans (NVG), Volga Germans versus all families, early-onset (NVG) versus all families, all early onset (Volga German + NVG) versus late-onset families and early-onset (NVG) versus late-onset families. The A test (Hodge et al. 1983) was used to test for heterogeneity within family groups (Volga Germans, early-onset NVG, and late-onset).

PCR Amplification and Direct Sequencing of APP exons 16 and 17. APP exon 17 was amplified as previously described using primers JS-23 (GTTGGGCAGAGAATATACTG) and JS-24 (CATGGAAGCACACTGATTCG) (Schellenberg et al, 1991a). The resulting amplified fragment was then digested with MboII to detect codon 693 mutations (Levy et al. 1990) and BclI to detect the codon 717 Val→Ile mutation (Goate et al. 1991). Exon 16 was amplified using primers APP-5 (GTTTTGGGTAGGCTTTG) and APP-3

(ACAGTAGTGGAAAGAGG) essentially as described for exon 17. Single stranded DNA for sequencing was generated by a second round of asymmetric PCR amplification essentially as described by Gyllensten and Erlich (1988). The asymmetric PCR product was used directly for dideoxy chain termination sequencing using Sequenase (United States Biochemical) and the instructions supplied by the manufacturer. For exon 17, the primers used for sequencing were either JS-23, JS-24, 17-152 (GTTCTTTGCAGAAGATGTGGG), or 17-293 (GCAGTCAAGTTTACCTACCTCC). For exon 16, the sequencing primer was either APP-3, APP-5, or KKAP5 (ATCCTGAGTCATGTCGG).

SSCP Analysis: Exon 17 was amplified using 200 ng of genomic DNA, 70 uM of each dNTP, 10 pmol of JS23, 10 pmol of JS24, 0.1 ul of α-^{32}P dCTP (3000 Ci/mmol), 10 mM Tris-HCl (pH 8.4), 50 mM KCl, 0.001% gelatin, 3.0 mM $MgCl_2$ and 1 unit Taq polymerase (Perkin-Elmer) in a final volume of 10 ul. Thermocycling conditions were 30 sec at 94°C, 30 sec at 60°C and 45 sec at 72° with 5 sec added to the 72° step for each of 30 cycles using a Thermal Cycler 9600 thermocycler (Perkin-Elmer). Five ul of the amplification product were digested with 10 units of StyI (or HinfI) in 20 ul according to the manufacturer's instructions. The digested sample was ethanol precipitated and dried and the pellet was resuspended in 5 ul of 0.1% SDS, 20 mM EDTA with vigorous vortexing. 1.5 ul of the sample was mixed with 9 ul of loading buffer composed of 95% formamide, 20 mM EDTA, 10 mM NaOH, 0.05 % xylene cyanol and 0.05 % bromphenol blue. The samples were resolved by electrophoresis using acrylamide gels prepared with

20 ml of 30% acrylamide mix (29:1), 10 ml of glycerol, 10 ml of 10 X TEB buffer, 60 ml of distilled water, 1.0 ml of 10% ammonium persulfate and 24 ul of TEMED. Samples were denatured at 96°C for 4 min, then put on ice. 2.5 ul of each sample was used for electrophoresis which was performed at 10 v/cm at room temperature for 28 hours. After drying, autoradiography was performed for 3 days at -70°C using an intensifying screen.

RESULTS

A total of 130 FAD kindreds were evaluated to determine whether mutations in the APP gene could be responsible for the disease. As an initial screen, all affected subjects from all families were evaluated for the FAD codon 717 Val→Ile mutation and the hereditary cerebral hemorrhage with amyloidosis-Dutch type (HCHWA-D) codon 693 Glu→Gln mutation. To detect these mutations, APP exon 17 was PCR amplified and the product digested with BclI (codon 717 Val→Ile) and MboII (codon 693 Glu→Gln). No examples of these mutations were detected in the FAD, spAD or normal control subjects tested (this study and Schellenberg et al. 1991). A single sample gave an altered MboII digest pattern but this sample did not have the HCHWA-D mutation (see below).

To evaluate whether APP mutations at other sites in the gene could cause AD, the larger families were evaluated for linkage of the APP gene region to FAD. Only families in which genotypes for 2 or more affected subjects could be obtained were used for

linkage analysis. The polymorphic marker used was D21S210, a highly informative dinucleotide repeat locus (heterozygosity = 83%, Warren et al. 1992a) which is tightly linked to APP (Z = 26.11, Θ_{max} = 0.01). This locus shows 1 recombination event with APP in 94 informative meioses in the C.E.P.H. pedigrees.

The results of the linkage analysis for D21S210 are given in Table 2. Families were divided into 3 groups; Volga Germans, early-onset (NVG), and late-onset families. The Volga German families were analyzed separately because this group is probably descended from a common ancestor and thus FAD in these families is likely to be genetically homogeneous (Bird et al. 1988). The group mean age-of-onset for the Volga German families is 57.9 yrs (SD = 8.63) years. The early-onset (NVG) families were those with family mean onsets $\leq$ 60 years (group mean = 46.4 yrs, SD = 6.20). For all 3 groups the summed LOD scores for close linkage of FAD to D21S210 were profoundly negative (Table 2) and obligate recombinants were noted in many of the families (Fig. 1). For the Volga German kindreds, linkage to the APP region was formally excluded beyond Θ = 0.10. Obligate recombinants were noted in 4 (HB, HD, KS and R) of the 7 families tested (Fig. 1). The KS family had a positive LOD score of Z = 1.04 (Θ_{max} = 0.15). However this family also showed an obligate recombinant between D21S210 and the disease. For the early-onset (NVG) families, obligate recombinants were observed in 7 (AM, HR-I, HR-XV, L, SNW, V and 603) of the 10 families tested. Linkage to the APP region was formally excluded beyond Θ = 0.20 (Table 2). Only the HR-I family showed a small positive LOD score (Z = 0.35 at Θ_{max} = 0.20) and this

family had an obligate recombinant between D21S210 and FAD. The highly negative LOD scores for D21S210 and the large number of obligate recombinants observed, particularly in the larger Volga German and the early-onset (NVG) kindreds, strongly indicates that few if any of these families have FAD mutations in the APP gene. For the Volga German families, the APP gene can be excluded for the group based on the combined data from the 7 families. Since D21S210 is most likely outside the APP gene, an obligate recombinant between FAD and D21S210 does not formally exclude the APP gene as the mutation site in a given family. However, since recombination between APP and D21S210 is a rare event, the highly negative LOD scores for linkage of FAD to D21S210 and the large number of recombinants observed provide strong exclusionary evidence for the APP gene in these families. Results from the heterogeneity tests also gave no evidence to support the hypotheses that a subset of families demonstrate linkage to APP; all tests for heterogeneity between and within family groups were non-significant (P ranged from 0.18 to 1.0).

For late-onset families, LOD scores for linkage of FAD to D21S210 were also negative (Table 2) and apparent obligate recombinants were noted in several families. Interpretation of the obligate recombinants in late-onset FAD is problematic. Since, unlike early-onset AD, late-onset AD is a relatively common disease, the possibility exists that in late-onset families "sporadic" cases are mixed with late-onset familial cases. Thus, although the negative LOD scores for the late-onset kindreds must be viewed with caution, these data suggest that APP is not the cause of AD in most of the late-onset

families studied. However, we cannot exclude the possibility that in a subset of the late-onset families, AD is caused by APP mutations.

Mutational Analysis of APP Exons 16 and 17. Since only a subset of the families studied were large enough to yield information by linkage analysis, and since interpretation of linkage results for the late-onset families is ambiguous, we looked for mutations in APP exons 16 and 17 by direct sequencing and SSCP analysis. The Volga German and early-onset (NVG) kindreds used in the linkage studies were screened to eliminate the possibility that recombinants between D21S210 and FAD were the result of a rare recombination event between APP and D21S210. SpAD subjects were evaluated to determine whether exon 17 contains susceptibility mutations which, because of reduced penetrance, age-censoring, or multilocus inheritance, do not result in obvious familial clustering of AD. Finally, elderly controls were examined to determine whether a protective variant exists and to identify changes in the coding region which do not alter the subject's susceptibility to develop AD. Exons 16 and 17 were chosen for screening for 2 reasons. First, the known FAD and HCHWA-D mutations are in exon 17. Second, AD may be the result of abnormal processing and/or degradation of APP at protease sites flanking the β-amyloid peptide sequence. Exons 16 and 17 encode the entire β-amyloid peptide sequence and 38 amino acids before and 23 amino acids after the β-amyloid peptide sequence.

In the initial screening for the HCHWA-D mutation, 1 sample from the SB family

gave a MboII digestion pattern consistent with the loss of the MboII site at codons 693/694 (Fig 2). Direct sequencing identified an A→G change at codon 693 resulting in a Glu→Gly substitution. The sample was from a subject with autopsy-documented AD with an age-of-onset of 56 years. Autopsy results showed neuritic plaques which stained positively with antibody to β-amyloid peptide, neurofibrillary tangles which stained with antibody to Tau, and moderate to severe amyloid associated with cortical and leptomeningeal blood vessels. The subject was from a small FAD kindred with a mean age-of-onset of 62 years (Table 1). An affected sib and 3 unaffected sibs did not have the 693_{Gly} allele (Fig 3). This variant was observed in 1 out of 127 FAD kindreds and was not observed in 141 spAD subjects or 277 normal controls (Table 3).

Direct Sequence Analysis. The coding regions and intron/exon boundaries of exons 16 and 17 were analyzed for mutations by direct sequencing of PCR-amplified genomic DNA. For exon 16, DNA from affected subjects from 24 early-onset and 51 late-onset families was sequenced. For exon 17, samples from affected subjects from 21 early and 20 late-onset FAD families and from 3 families with other familial neurodegenerative disorders were sequenced for the entire exon and intron/exon boundaries. For an additional 22 late-onset families, the partial sequence for the coding region including codon 717 was obtained. This direct sequencing approach identified 2 variants. One was in the coding region of exon 17 at codon 708 in a sample from a family with Parkinson's disease, depression, tremor and multiple sclerosis (LIT family). The change did not alter the amino acid at this position. This variant was also detected in 1 spAD subject.

No variants were observed in exon 16 coding sequences.

The second variant identified by direct sequencing was a 6 bp deletion in intron 17. This deletion was identified in a total of 3 affected subjects in the HEU and FZG families (Fig 3) but was not present in 1 unaffected sib from the FZG family. When other FAD, spAD and controls were screened for this variant, 1 other example was identified. This subject was a 92 year old normal control who was remarkably healthy and living independently in her own home. She had a 94 year old sister with probable AD (onset 83 yrs) who did not have the deletion.

SSCP Analysis. Samples from 30 early-onset FAD kindreds, 98 late-onset FAD kindreds, 153 spAD subjects and 195 normal controls were screened by SSCP for mutations in exon 17 and flanking regions. Two variants were identified by this method. One was a C→T change 114 bp before the beginning of exon 17 (Fig. 2c). This allele was observed in 3 subjects (1 spAD and 2 non-demented controls). Interestingly, all 3 subjects with this allele were African-Americans. Since only 13 African American subjects were tested, this variant may be a polymorphism in individuals of African descent. The second variant detected by SSCP is a T→C change in intron 16, 56 bp before the start of exon 17 (Fig 2d). This allele was observed in a single spAD subject.

Detection of APP Mutations and Variants. All of the variants described here and the codon 717 Val→Ile FAD mutation can be detected by SSCP (Fig 4). The codon 693

Glu→Gly change results in the loss of an MboII site. This change can be easily detected after electrophoresis by ethidium bromide staining. However, this method does not distinguish between the 693_{Gly} allele reported here and the HCHWA-D 693_{Gln} allele. The intron 16 C→T allele results in the loss of an MnlI restriction site and can be detected by the same procedure. The 6 bp deletion results in the loss of a DdeI site. This variant can be detected by end-labeling primer JS24, PCR amplification of a 417 bp fragment containing exon 17, digestion of the product with either DdeI or AluI, and resolving the resulting fragments by electrophoresis. The rare codon 708_{Gly} allele and the intron 16 T→C alleles do not alter restriction sites and must be detected either by SSCP or by direct sequencing.

Linkage Analysis of the centromeric region of chromosome 21. Although much of the positive data for linkage of chromosome 21 markers to FAD was the result of families with APP mutations (Schellenberg et al. 1991a), analysis of a French-Italian FAD kindred provides at least suggestive evidence that a second FAD locus, centromeric to APP, could exist on chromosome 21. This family shows obligate recombinants between APP and FAD (Tanzi et al. 1987) yet yields positive LOD scores for markers centromeric to APP. A peak LOD score of 2.94 ($\Theta = 0.15$) was reported for D21S52 (Tanzi et al. 1991). The order of markers in this region is APP, D21S1/S11, D21S52, D21S13, centromere (Tanzi et al. 1992). Others have suggested that there may be a locus centromeric to D21S13 (St George-Hyslop et al. 1991). To determine if there was any evidence for a second chromosome 21 locus in our family collection, we evaluated linkage of D21S1/S11,

D21S13, and D21S215 to FAD. D21S215 is a pericentric dinucleotide repeat marker which is centromeric to D21S13 (Warren et al. 1992). The LOD scores obtained were substantially negative for all family groups tested. Multipoint analysis of this region for the Volga German kindreds and a subset of the early-onset (NVG) kindreds were also negative for the entire region (Fig 5). The μ-test for heterogeneity, when applied to the multipoint data (Fig. 5), also gave no evidence of heterogeneity between the Volga German and early-onset (NVG) families (χ_1^2 = 1.16, P = 0.28), which is consistent with evidence against linkage to this region. Only 2 families gave LOD scores of > 1; the SNW family which yielded a LOD score of 1.26 at Θ_{max} = 0.001 for D21S1/S11 and the HD family which gave a LOD score of 1.17 at Θ_{max} = 0.001 for D21S215. Both families gave negative LOD scores or very small positive LOD scores (Z < 0.03) at all recombination fractions for the other markers in Table 4 and for D21S210 (Table 2). These data do not support the hypothesis that there is a second chromosome 21 FAD locus in the families reported here. However, since FAD is clearly genetically heterogeneous, the possibility exists that in a limited number of families, AD is caused by a locus in this region.

DISCUSSION

Mutations in the APP gene clearly are responsible for FAD in a small number of early-onset families (Goate et al. 1991; Murrell et al. 1991; Chartier-Harlin et al. 1991b). The above work was undertaken to determine whether FAD in other early-onset

kindreds is caused by APP gene mutations and to determine whether APP mutations could be responsible for late-onset and spAD forms of the disease. Linkage analysis of the APP gene region was performed using a newly developed short tandem repeat polymorphism (D21S210) which is tightly linked to APP. This marker is much more informative for the APP gene region compared to the 2 allele RFLPs used in previous studies. The results indicate that in most if not all of the 10 early-onset (NVG) families studied, FAD is not caused by mutations in the APP gene. Seven of the 10 early-onset (NVG) kindreds showed obligate recombinants between D21S120 and FAD. The remaining 3 families yielded only negative LOD scores. Similarly, for the Volga German kindreds, multiple recombinants were observed between D21S210 and FAD and LOD scores for linkage of FAD to D21S210 were extremely negative. Heterogeneity tests gave no evidence for a mixture of linked and unlinked families. Thus, an APP mutation is not responsible for FAD in the Volga German families.

FAD families which did not have a sufficient number of living subjects which could be sampled and thus were not suitable for linkage, were analyzed by direct sequencing and SSCP. Often only a single subject was available for sampling with other affected subjects in the family deceased. We detected no examples of the known FAD or HCHWA-D mutations or other variants that could be unambiguously described as FAD mutations. This work demonstrates that FAD mutations in exons 16 and 17, including the known codon 717 mutations, are rare in early-onset kindreds and have not been found in late-onset or spAD subjects. We also did not observe a variant present in only

controls and thus found no evidence for a protective allele. The possibility still exists, however, that mutations in other APP exons could be responsible for some AD subjects, particularly in the late-onset and spAD groups. Although no late-onset FAD mutation has been identified to date, only exons 16 and 17 have been exhaustively examined.

Five DNA sequence variants in and around APP exon 17 were identified in this study. The most interesting variant was the codon 693 Gly allele found in the SB family. At this same codon, a Gln→Glu change results in HCHWA-D, an autosomal dominant disorder. Although HCHWA-D shares some neuropathologic features with AD, it is neuropathologically and clinically distinct disease. In HCHWA-D, the β-amyloid peptide is found deposited as amyloid at cerebral vascular sites but AD-type neuritic plaques with amyloid and neurofibrillary tangles are not typically found. Clinically HCHWA-D patients suffer from cerebral hemorrhage and sometimes a multi-infarct type dementia (Haan et al. 1990a,b) which is not the predominant clinical feature of AD. The HCHWA-D mutation is close to the constitutive APP cleavage site (codon 687, Sisodia et al. 1990; Esch et al. 1990). Cleavage at this site results in the release of a soluble form of APP (nexin-II) and in the destruction of the amyloidgenic fragment of APP. Conceivably, alterations at codon 693 could affect the efficiency of this cleavage and lead to accelerated deposition of the β-amyloid peptide. The 693 Gly allele observed in the SB family did not co-segregate with FAD in the small number of samples available from that family (Fig 3), since only 1 of the 2 affected subjects had the 693_{Gly} allele. However, this rare variant could conceivably be a predisposing allele in a family where

other genes are also involved in producing AD. The alternative explanation is that this 693_{Gly} allele is simply a rare variant unrelated to FAD. Additional FAD and control subjects will need to be tested to determine the significance of this allele.

The 6 bp deletion variant in intron 17 does co-segregate with FAD in 3 meiosis in 2 FAD families (Fig 3). However, since this deletion is outside of the coding regions of the gene, is distant from the exon/intron boundaries, and was also found in a cognitively intact control, this variant is probably not an FAD mutation. A less likely possibility is that this deletion is an incompletely penetrant mutation which alters an APP regulatory site and predisposes an individual to FAD.

The codon 708 variant is neutral and does not alter the amino acid at this position. This single base change is outside the postulated translational regulatory stem-loop sequence described by Tanzi and Hyman (1991) and is thus probably unrelated to AD. The 2 single base variants observed in intron 16 are also probably unrelated to AD since both are in non-coding regions and are far from the intron/exon splice sites.

The conclusion from the work presented here and earlier linkage studies of FAD and the APP gene (Van Broeckhoven et al. 1987; Tanzi et al. 1987b; Schellenberg et al. 1988) indicate that in most early-onset FAD families, mutations at the APP gene are not responsible for the disease. Both linkage and mutational analysis (see also Crawford et al. 1991) also indicate that the known FAD APP codon 717 mutations are rare among

early and late-onset kindreds. Similarly, analysis of chromosome 21 markers centromeric to APP also indicate that FAD in the Volga German families and the majority of the early-onset kindreds studied are not the result of a second centromeric locus. Thus, at least 1 other FAD locus exists and remains to be identified. Pericak-Vance et al. (1991) have reported evidence for an FAD gene on chromosome 19 in late-onset families. The early-onset kindreds described above do not show evidence of linkage to this region of chromosome 19 (GD Schellenberg and HT Orr, unpublished data) and thus the location of the gene(s) responsible for most early-onset FAD remains to be identified. We are attempting to identify other FAD loci by linkage analysis using markers scanning the genome. Because genetic heterogeneity can confound linkage results, it is particularly important to screen families to be used for future linkage studies to make sure that families with APP mutations are not included.

Amyloid depositions consisting of primarily the β-amyloid peptide in the central nervous system have long been recognized as the distinctive neuropathologic hallmarks of AD. However, it has been difficult to establish whether formation of the β-amyloid peptide and/or deposition is a primary event in the pathogenic process or the result of cell injury. The APP codon 717 FAD mutations are perhaps the best evidence implicating β-amyloid formation as the key event in the initiation of the neurodegenerative events leading to AD. The actual steps in APP processing, the relevant proteases, and the potential protease inhibitors of alternative modes of processing remain to be unambiguously identified. Characterization of other FAD loci

may provide additional information on the different events leading to the production of the β-amyloid peptide and its associated neurotoxicity.

ACKNOWLEDGEMENTS

Supported by NIA grants AG05136 (Alzheimer's Disease Research Center at the University of Washington), AG08017 (Alzheimer's Disease Center Core), AG06781, and AG07584, NIMH grant R01MH 43240 (HTO,JAW, LLH), and NIH grants HD24605 and HG00373. Also supported by GM 15253 Veterans Administration Research Funds (TDB), the American Health Assistance Foundation (TDB, GDS, JRK, SMP), the Alzheimer's Society of British Columbia (ADS), the Washington Friends of Alzheimer's Research Program (KK), and the Alzheimer's Disease and Related Disorders Association (JRK, MM, AW, SEA, HP). We thank Gertrude Shiely and Katherine Kiech for gifts to the University of Minnesota. TDB, GDS, EMW, and GMM are in the University of Washington Alzheimer's Disease Research Center. MJB, HP and JK are in the Alzheimer's Disease Research Center at the Oregon Health Sciences University. We thank D. Nochlan and M. Sumi for neuropathologic characterization of many of the families, Elisa Alonso (National Institute of Neurology and Neurosurgery) for help in sampling the HR-XV family, Elaine Loomis and Siri Ito for technical help, and H. Lipe, N. Chinn, J. Thompson and M. Pfanschmidt for obtaining blood samples and medical records. We also thank the ADRC clinical core, Group Health Cooperative of Puget Sound, and Dr. R. Tatham (Homewood Hospital Guelph, Ontario) for identification and

characterization of some of the subjects. AD Roses kindly provided cell lines from family 372 containing the APP codon 717 mutation and some of the cell lines for family 603. Informed consent was obtained from each subject or next-of-kin with approval of the University of Washington Human Subjects Review Committee or the University of Western Ontario Ethics Committee.

REFERENCES

Bird TD, Lampe TH, Nemens EJ, Miner GW, Sumi SM, Schellenberg GD (1988) Familial Alzheimer's disease in American descendants of the Volga Germans: probable genetic founder effect. Ann Neurol 23:25-31

Bird TD, Sumi SM, Nemens EJ, Nochlin D, Schellenberg GD, Lampe TH, Sadovnick A, et al (1989) Phenotypic heterogeneity in familial Alzheimer's disease: a study of 24 kindreds. Ann Neurol 25:12-25

Chartier-Harlin M-C, Crawford F, Hamandi K, Mullan M, Goate A, Hardy J, Backhovens H, et al (1991a) Screening for the β-amyloid precursor mutation (APP717: Val→Ile) in extended pedigrees with early-onset Alzheimer's disease. Neurosci Lett 129:134-135

Chartier-Harlin M-C, Crawford F, Houlden H, Warren A, Hughes D, Fidani L, Goate A, Rossor M, et al (1991b) Early-onset Alzheimer's disease caused by mutation at codon

717 of the β-amyloid precursor protein gene. Nature 353:844-846

Cook R, Ward BE, Austin JH, (1979) Studies in Aging of the Brain, IV. Familial Alzheimer's Disease: Relation to Transmissible Dementia, Aneuploidy and Microtubular Defects. Neurology 29:1402-1412

Crawford F, Hardy J, Mullan M, Goate A, Hughes D, Fidani L, Roques P, Rossor M, Chartier-Harlin M-C (1991) Sequencing of exons 16 and 17 of the β-amyloid precursor protein gene in 14 families with early onser Alzheimer's disease fails to reveal mutations in the β-amyloid sequence. Neurosci Lett 133:1-2

De Sauvage F, Octave J-N (1989) A novel mRNA of the A4 amyloid precursor gene coding for a possible secreted protein. Science 245:651-653

Esch FS, Keim PS, Beattie EC, Blacher RW, Culwell AR, Oltersdorf T, McClure D, Ward PJ (1990) Cleavage of amyloid β peptide during constitutive processing of its precursor. Science 248:1122-1124

Glenner GG, Wong CW (1984) Alzheimer's disease and Down's syndrome: sharing of a unique cerebrovascular amyloid fibril protein. Biochem Biophys Res Comm 122:1131-1135

Goate AM, Chartier-Harlin CM, Mullan M, Brown J, Crawford F, Fidani L, Giuffra L et al (1991) Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature 349:704-706

Gou Z, Sharma V, Litt M. (1990) Dinucleotide repeat polymorphism at the D21S13E locus. Nucl Acids Res 18:4967

Goudsmit J, White BJ, Weitkamp LR, Keats, BJB, Morrow CH Gajdusek DC (1981) Familial Alzheimer's Disease in Two Kindreds of the Same Geographic and Ethnic Origin. J. Neurol Sci 49:79-89

Gyllensten UB, Erlich HA (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. Proc Natl Acad Sci USA 85:7652-7656

Haan J, Algra PR, Roos RAC (1990a) Hereditary cerebral hemorrhage with amylodosis-Dutch type. Clinical and comouted tomographic analysis of 24 cases. Arch Neurol 47:649-653

Haan J, Lanser JBK, Zijderveld I van der Does IGF, Roos RAC (1990b) Dementia in hereditary cerebral hemorrhage with amylodosis-Dutch type. Arch Neurol 47:965-967

Heston LL, Orr HT, Rich SS, White JA (1991) Linkage of an Alzheimer disease susceptibility locus to markers on human chromosome 21. Am J Med Genet 40:449-453

Hodge SE, Anderson CE, Neiswanger K, Sparkes RS Rimoin DL (1983) The search for heterogeneity in insulin-dependent diabetes mellitus (IDDM): Linkage studies, two-locus models, and genetic heterogeneity. Am J Hum Genet 35:1139-1155

Kang J, Lemaire H-G, Unterbeck A, Salbaum JM, Masters CL, Grzeschik K-H, Multhaup G et al (1987) The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. Nature 325:733-736

Kitaguchi N, Takahashi Y, Tokushima Y, Shiojiri S, Ito H (1988) Novel precursor of Alzheimer's disease amyloid protein shows protease inhibitory activity. Nature 311:530-532

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

Lemaire HG, Salbaum JM, Multhaup G, Kang J, Bayney RM, Unterbeck A, Beyreuther K, Muller-Hill B (1989) The PreA4$_{695}$ precursor protein of Alzheimer's disease A4 amyloid is encoded by 16 exons. Nucl Acids Res 17:517-522

Levy E, Carman MD, Fernandez-Madrid IJ, Power MD, Lieberburg I, Van Duinen SG, Bots GThAM, et al (1990) Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrhage, Dutch type. Science 248:1124-1126

Morton NE (1956) The deletion and estimation of linkage between the genes for elliptocytosis and the Rh blood type. Am J Hum Genet 8:80-96

Naruse S, Igarashi S, Aoki K, Kaneko K, Iihara K, Miyatake T, Kobayashi H, et al (1991) Mis-sense mutation Val-Ile in exon 17 of amyloid precursor protein gene in Japanese familial Alzheimer's disease. Lancet 337:978-979

Pericak-Vance MA, Bebout JL, Gaskell PC, Yamaoka LA, Hung W-Y, Alberts MJ, Walker AP, et al (1991) Linkage studies in familial Alzheimer's disease: evidence for chromosome 19 linkage. Am J Hum Genet 48:1034-1050

Ponte P, Gonzalez-DeWhitt P, Schilling J, Miller J, Hsu D, Greenberg B, Davies K, et al (1988) A new A4 amyloid mRNA contains a domain homologous to serine protease inhibitors. Nature 331:525-527

Schellenberg GD, Bird TD, Wijsman EM, Moore DK, Boehnke M, Bryant EM, Lampe TH, Sumi SM, Deeb SM, Beyreuther K, Martin GM (1988) Absence of linkage of chromosome 21 q21 markers to familial Alzheimer's disease in autopsy-documented

pedigrees. Science 241:1507-1510

Schellenberg GD, Anderson L-J, O'dahl S, Wijsman EM, Sadovnick AD, Ball MJ, Larson EB, et al (1991a) APP_{717}, APP_{693}, and PRIP gene mutations are rare in Alzheimer disease. Am J Hum Genet 49:511-517

Schellenberg GD, Pericak-Vance MA, Wijsman EM, Moore DK, Gaskell PC, Yamaoka LA, Bebout JL et al (1991b) Linkage analysis of familial Alzheimer's disease using chromosome 21 markers. Am J Hum Genet 48:563-583

Schellenberg GD, Boehnke M, Wijsman EM, Moore DK, Martin GM, Bird TD (1992) Genetic association and linkage analysis of the Apolipoprotein CII locus and familial Alzheimer's disease. Ann Neurol 31:223-227

Sharma V, Litt M (1992) Tetranucleotide repeat polymorphism at the D21S11 locus. Hum Molec Genet (in press)

Sisodia SS, Koo EH, Beyreuther K, Unterbeck A, Price DL (1990) Evidence that the β-amyloid protein in Alzheimer's disease is not derived by normal processing. Science 248:492-495

St George-Hyslop PH, Haines JL, Farrer LA, Polinsky R, Van Broeckhoven C, Goate

A, Crapper DR et al (1990) Genetic linkage studies suggest that Alzheimer's disease is not a single homogeneous disorder. Nature 347:194-197

Stinissen P, Van Broeckhoven C (1991) A new (CA)n repeat polymorphism at the D21S13E locus. Nucl Acids Res 19:5089

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

Tanzi RE, St George-Hyslop PH, Haines JL, Polinsky RJ, Nee L, Foncins J-F, Neve RL, McClatchey AI, Conneally PM, Gusella JF (1987) The genetic defect in familial Alzheimer's disease is not tightly linked to the amyloid beta-protein gene. Nature 329:156-157

Tanzi RE, McClatchey AI, Lamperti ED, Villa-Komaroff L, Gusella J, Neve RL (1988) Protease inhibitor domain encoded by an amyloid protein precursor mRNA associated with Alzheimer's disease. Nature 331:528-530

Tanzi RE, Hyman BT (1991) Alzheimer's mutation. Nature 350:564

Tanzi RE, St George Hyslop P, and Gusella JF (1991) Molecular genetics of Alzheimer

disease amyloid. J Biol Chem 266:20579-20582

Tanzi RE, Watkins PC, Stewart GD, Wexler NS, Gusella JF, Haines JL (1992) A genetic linkage map of human chromosome 21: analysis of recombination as a function of sex and age. Am J Hum Genet 50:551-558

Valencia LG, Alonso ME, Figeroa HH, Escobar A (1986) Enfermedad de Alzheimer Presentacion de 7 casos en 3 familias. La Rev Invest Clin (Mex) 38:261-267

Van Broeckhoven C, Genthe AM, Vandenberghe A, Horsthemke B, Backhovens H, Raeymaekers P, Van Hul W, et al (1987) Failure of familial Alzheimer's disease to segregate with the A-4 amyloid gene in several European families. Nature 329:153-155

Warren AC, McInnis MG, Kalaitsidaki M, Cox TK, Chakravarti A, Antonarakis SE (1992a) D21S210 is completely linked to the β-amyloid precursor (APP) gene, and is identified by a highly polymorphic (GT)n Marker. (submittted)

Warren AC, Petersen MB, Van Hul W, McInnis MG, Van Broeckhoven C, Cox TK, Chakravarti A, Antonarakis SE (1992b) Dinucleotide repeat polymorphisms on human chromosome 21. (submitted)

Williamson R, Bowcock A, Kidd K, Pearson P, Schmidtke J, Ceverha P, Chippenfield 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

Yoshida K, Miki T, Katsuya T, Ogihara T, Sakaki Y (1991) The ^{717}Val→Ile substitution in amyloid precursor protein is associated with familial Alzheimer's disease regardless of ethnic groups. Biochem Biophys Res Comm 178:1141-1146

Table 1. Characteristics of FAD families not previously described.

Family	Total Number affecteds	Total number autopsies	Number[a] affecteds tested	Number of subjects sampled	Mean age-of-onset ± SD (n)
B	1 (2)[b]	0	1	1	
BAW	2	0	1	1	79 ± 3.5 (2)
BOB[c]	5 (2)	2	3	6	67 ± 9.3 (2)
BOS	3	1	1	2	70 ± 0.8 (3)
BWH	3 (1)	0	3	3	75 ± 3.4 (3)
FL	1	0	1	1	70 (1)
HR-IX	7	3[d]	5	14	72 ± 1.9 (6)
HR-XI	5	2	4	23	64 ± 5.4 (5)
HR-XVc	12	1e	4	15	42 ± 3.9 (6)
HSP	1 (4)	0	1	1	75 (1)
KAM[c]	6 (5)	0	4	9	72 ± 8.2 (5)
LJT[c]	6	1	4	7	74 ± 6.5 (4)
NB	1	0	1	1	73 (1)
ORAS[c]	5	1	1	2	70 ± 1.5 (2)
ORCW[c]	8	1	6	8	72 ± 7.3 (7)
ORDF[c]	3	0	2	2	71 ± 0.9 (3)
OREB[c]	5 (1)	0	2	3	73 ± 0.5 (2)
ORGK[c]	6 (1)	2	2	6	66 ± 8.3 (5)
ORGL[c]	3	0	2	2	66 ± 2.5 (2)
ORHM[c]	5	0	3	5	78 ± 5.4 (3)

Table 1. (cont)

Family	Total Number affecteds	Total number autopsies	Number[a] affecteds tested	Number of subjects sampled	Mean age-of-onset ± SD (n)
ORHW[c]	3	0	2	4	69 ± 8.2 (3)
ORJM[c]	5	1	1	3	74 ± 4.3 (3)
ORJV[c]	5	1	2	2	77 ± 2.5 (3)
ORKD[c]	3	0	2	7	78 ± 1.4 (3)
ORLA[c]	5	1	2	2	74 ± 1.2 (5)
ORLH[c]	7	1	4	6	75 ± 1.5 (5)
ORLR[c]	4	1	2	2	68 ± 0.5 (3)
ORMK[c]	2	0	2	2	68 ± 1.5 (2)
ORPH[c]	7	1	2	3	69 ± 8.1 (4)
ORPK[c]	7	1	3	6	78 ± 2.9 (3)
ORRH[c]	3	0	2	3	72 ± 6.0 (2)
ORSK[c]	7	1	3	5	75 ± 0.0 (2)
ORSS[c]	5	0	2	5	79 ± 4.5 (4)
ORTT[c]	4	0	2	2	69 ± 0.0 (1)
ORVT[c]	2 (1)	0	1	2	55 ± 0.0 (1)
ORWW[c]	8	0	3	4	67 ± 7.0 (4)
RBR	4 (1)	1	1	2	76 ± 6.0 (2)
RC	5	0	2	11	
RPN	3 (1)	0	1	1	62 ± 3.3 (3)
S	1 (5)	0	1	1	

Table 1. (cont)

Family	Total Number affecteds	Total number autopsies	Number[a] affecteds tested	Number of subjects sampled	Mean age-of-onset ± SD (n)
SB[c]	5	1	2	5	62 ± 7.3 (5)
SD80	5	3[e]	1	1	68 ± 1.4 (4)
SEH	3	0	2	3	63 ± 5.7 (3)
SLE	1 (5)	0	1	1	69 (1)
SMM	4	0	2	4	71 ± 6.0 (2)
SSW	6	2	1	2	70 ± 3.4 (3)
STM	1 (1)	0	1	1	
UBH	(5)	0	1	1	
WKW	3 (1)	0	1	1	71 ± 10.4 (3)
WFLc	6	2	2	12	64 ± 7.6 (6)

[a] All affected subjects for whom DNA was available were tested for the APP codon 717 and 693 mutations.

[b] Values in parenthesis indicate number of family members with the diagnosis of possible AD or demented by family history.

[c] Families used for linkage analysis.

d Includes 1 unaffected subject autopsied.

[e] Brain biopsy

[f] Includes 2 unaffected subjects autopsied.

Table 2. LOD scores for linkage of FAD to D21S210.

Family	Recombination Fraction (Θ)						
	0.001	0.05	0.10	0.15	0.20	0.30	0.40
Volga Germans							
E	-0.12	-0.09	-0.06	-0.04	-0.02	0.00	0.00
H	0.27	0.23	0.19	0.15	0.12	0.06	0.01
HB	-7.70	-3.10	-1.97	-1.32	-0.88	-0.35	-0.08
HD	-4.67	-1.90	-1.24	-0.86	-0.59	-0.26	-0.08
KS	0.02	0.84	1.01	1.04	0.99	0.76	0.39
R	-2.53	-1.03	-0.42	-0.10	0.07	0.18	0.11
W	-0.39	-0.24	-0.15	-0.09	-0.06	-0.02	0.00
Totals	-15.12	-5.29	-2.64	-1.22	-0.37	0.37	0.35
Early-Onset (NVG)							
AM	-1.59	-1.11	-0.70	-0.45	-0.29	-0.11	-0.02
HR-I	-1.68	-0.43	0.06	0.27	0.35	0.32	0.17
HR-X	-0.20	-0.15	-0.11	-0.09	-0.06	-0.03	-0.01
HR-XIII	-1.52	-0.51	-0.26	-0.13	-0.05	0.01	0.02
HR-XV	-2.62	-1.59	-1.11	-0.80	-0.56	-0.25	-0.07
KG	-0.66	-0.57	-0.47	-0.38	-0.29	-0.13	-0.03
L	-9.81	-3.94	-2.58	-1.79	-1.25	-0.57	-0.20
603	-1.03	-0.33	-0.14	-0.05	-0.02	-0.03	-0.05

Table 2. (cont.)

Family	Recombination Fraction (Θ) 0.001	0.05	0.10	0.15	0.20	0.30	0.40
SNW	-2.26	-0.92	-0.50	-0.28	-0.15	-0.02	0.02
V	-2.43	-0.94	-0.61	-0.41	-0.27	-0.11	-0.04
Total	-26.46	-11.66	-7.04	-4.39	-2.69	-0.83	-0.12
Late-onset (Autopsied)	-11.53	-5.77	-3.40	-1.98	-1.08	-0.16	0.11
	(-3.72)	(-1.79)	(-0.96)	(-0.50)	(-0.22)	(0.01)	(0.07)
Late-onset (All)	-18.23	-9.27	-5.62	-3.42	-1.99	-0.47	0.05
	(-7.01)	(-3.08)	(-1.58)	(-0.78)	(-0.33)	(0.04)	(0.10)

a LOD scores were calculated using age-corrected penetrance and a FAD gene frequency of 0.01. LOD scores in parenthesis were calculated at 1% penetrance and an FAD gene frequency of 0.01. All Volga German kindreds and early-onset families have at least 1 case of autopsy-documented AD. Autopsy-documented late-onset families include those indicated in Table 1 and families BCR, BE, BEH, BL, BMH, CFD, CK, CSF, EL, FH, HRA9, JR, MI, MMM, P, PH, QDH, RR, WLA (Schellenberg et al 1991a, 1991b), HR-XII, and HRA9 (Heston et al. 1991). Late-onset families lacking autopsy documentation not listed in Table 1 are FF, HTB and LQ (Schellenberg et al. 1991a).

Table 3. Frequencies of APP Gene Variants and Mutations[a]

Variant/mutation	Early-onset FAD Families[b]	Late-onset FAD Families	spAD Subjects	Age-matched Normal Subjects[c]
	(subjects with variant allele/total number subjects tested)			
717 Ile[d]	0/30	0/98	0/153	0/307
717 Phe or Gly	0/21	0/42	n.d.	n.d.
693 Gly	0/29	1/98	0/141	0/289
708 Gly→Gly	0/29	0/98	1/149	0/194
intron 16 T→C	0/29	0/98	1/149	0/194
intron 16 C→T	0/29	0/98	1/149	2/194
intron 17 deletion	0/29	2/98	0/153	1/207

[a] For early-onset FAD and late-onset FAD families, results are given as families with variant/total families sampled. For spAD and controls, results are given as number of subjects with variant or mutation/total subjects screened. The APP codon 693 Glu→Gly variant was screened for by MboII digestion and/or SSCP (Fig. 4). The APP codon 717 Val→Ile mutation was screened for by direct sequencing, digestion with DdeI, or by SSCP or a combination of these methods. The codon 717 Val→Gly and Val→Phe mutations were screened for by direct sequencing. The 2 intron 16 variants were screened for by SSCP. The intron 17 deletion was screened for by digestion with Alu or by SSCP as shown in Fig. 4. "n.d." indicates not determined.

[b] Includes 4 Volga German families.

[c] Includes 12 subjects originally diagnosed as "probable AD" but who were re-classified on follow up as either "possible AD" or as non-AD. For the 717 Ile and 693 Gly, the controls include 100 and 82 young normal controls, respectively.

[d] Includes data from the previously published study (Schellenberg et al. 1991a).

Table 4. LOD scores for linkage of FAD to Chromosome 21 Markers

Marker	Group	Recombination Fraction (Θ)						
		0.001	0.05	0.10	0.15	0.20	0.30	0.40
D21S215	Volga German	-7.45	-2.20	-0.64	0.12	0.50	0.65	0.38
		(-0.70)	(-0.19)	(0.02)	(0.12)	(0.16)	(0.16)	(0.10)
	Early-onset (NVG)	-13.28	-5.06	-2.76	-1.52	-0.78	-0.10	0.06
		(-3.88)	(-2.05)	(-1.14)	(-0.69)	(-0.41)	(-0.11)	(0.00)
	Late-onset (autopsied)	-7.86	-3.46	-2.06	-1.24	-0.71	-0.17	0.02
		(-3.74)	(-1.40)	(-0.61)	(-0.22)	(-0.03)	(0.08)	(0.04)
	Late-onset (All)	-9.91	-4.97	-3.16	-2.03	-1.26	-0.39	-0.01
		(-4.23)	(-1.80)	(-0.91)	(-0.44)	(-0.18)	(0.02)	(0.03)
D21S13	Volga German	-9.19	-4.60	-2.86	-1.78	-1.08	-0.32	-0.07
		(-1.80)	(-1.09)	(-0.66)	(-0.40)	(-0.24)	(-0.08)	(-0.03)
	Early-onset (NVG)	-3.93	-1.95	-1.26	-0.84	-0.55	-0.22	-0.06
		(-1.37)	(-0.84)	(-0.57)	(-0.39)	(-0.28)	(-0.13)	(-0.05)
	Late-onset (Autopsied)	-12.25	-6.72	-4.40	-2.96	-1.96	-0.75	-0.17
		(-5.12)	(-2.89)	(-1.85)	(-1.22)	(-0.81)	(-0.31)	(-0.08)
	Late-onset (All)	-14.09	-7.06	-4.36	-2.75	-1.69	-0.54	-0.09
		(-5.55)	(-2.96)	(-1.78)	(-1.10)	(-0.69)	(-0.24)	(-0.05)
D21S1/S11	Volga German	-12.00	-6.12	-3.94	-2.60	-1.71	-0.66	-0.17
		(-1.74)	(-0.93)	(-0.53)	(-0.30)	(-0.17)	(-0.04)	(0.00)
	Early-onset (NVG)	-4.92	-1.54	-0.56	-0.08	0.15	0.24	0.12
		(-1.01)	(-0.46)	(-0.20)	(-0.05)	(0.02)	(0.05)	(0.03)

Table 4. (cont.)

Marker	Group	Recombination Fraction (Θ) 0.001	0.05	0.10	0.15	0.20	0.30	0.40
D21S1/S11	Late-onset (Autopsied)	-7.46	-2.58	-1.58	-0.89	-0.49	-0.09	0.02
		(-4.66)	(-2.32)	(-1.37)	(-0.83)	(-0.49)	(-0.14)	(-0.02)
	Late-onset (All)	-9.04	-3.53	-1.95	-1.08	-0.58	-0.10	0.02
		(-4.43)	(-1.74)	(-0.77)	(-0.30)	(-0.07)	(0.08)	(0.04)

a LOD scores were calculated using age-corrected penetrance and a FAD gene frequency of 0.01. Values in parenthesis were calculated using 1% penetrance and an FAD frequency of 0.01.

FIGURE LEGENDS

Fig 1. APP and D21S210 Genotypes for Volga German and early-onset FAD kindreds. Pedigrees illustrated are those which show obligate recombination events between FAD and either D21S210 or APP. Partial pedigrees are shown which include sibships showing obligate recombination events between FAD and D21S210 (alleles are 50-80 and are the actual fragment size in bp minus 100) and/or APP BglII genotypes (alleles are 6.9 and 9.3). For families AM, L, SNW and 603, for each sibship shown, all known subjects are included. For families HB, HD, KS, R, HRI, and HR-XV, only subjects relevant to recombination between FAD and D21S210 and/or APP loci are shown. For more complete pedigrees see the following references: Bird et al. (1989) for L, R (see also Cook et al. 1979), HB, V, KS, and HD; Schellenberg et al. (1991b) for 603 and AM; Heston et al. (1991) for HR-I; Goudsmit et al. (1981) for SNW; and Valencia et al. (1986) for HR-XV (family 3 in original reference). Circles are females, squares are males, a filled in circle or square represents affected subjects, and a slash indicates the subject is dead. A stippled symbol indicates that the AD status of the subject is unknown. The number above and to the left of subjects is the age-of-onset for affected subjects, the current age for living subjects or the age at death for unaffected subjects. For affected subjects, if the age-of-onset is not known, the age of death is shown preceded by "D". An "A" below and to the right indicates the subject was autopsied. Several non-demented subjects in the L family were autopsied and found to be neuropathologically normal. Genotypes in parenthesis were deduced from spouse and

offspring genotypes.

Fig 2. DNA sequences of APP variants. In A, a partial sequence for the coding region of exons 16 and 17 is given. The numbers below the amino acids are the position of that residue in the β-amyloid peptide. The transmembrane region of the protein is underlined. The sequence in B is from the intron following exon 17 and begins at bp 53 of this intron (Lemaire et al. 1989). The exon numbering system used here is based on the 770-splice form of the APP protein and exon 17 is equivalent to exon 15 in Lemaire et al. (1989). The location of the 6 bp deletion is ambiguous by 1 base and both possible sites are underlined. The sequences in C and D are from intron 16 single base variants are indicated above each sequence. These sequences begin 129 bp and 69 bp, respectively, from the beginning of exon 17.

Fig 3. Segregation of APP variants in FAD families. Symbols are as in Fig 1. A "+" below an individual indicates the presence of the 6 bp deletion (FZG and HEU families) or the exon 17 codon 693 Glu→Gly variant (SB family). A "-" indicates the normal allele.

Fig 4. Detection of APP variants. A, detection of the codon 693 Glu→Gly variant by MboII digestion. The 417 bp PCR product containing exon 17 was digested with MboII. The fragments were resolved by electrophoresis using a non-denaturing polyacrylamide gel and visualized by ethidium bromide staining. Samples are 1 and 9, HaeIII ΦX174

(BRL) size standard; 2 and 8, Marker V (Boehringer Mannhein) size standard; and 3-6, SB family sibs in the order shown in Fig 3. B, detection of intron 16 C→T variant by MnlI digestion. The same 417 PCR product was digested with MnlI and the fragments resolved as in A. Lanes are: 1 and 10, 123 bp ladder (BRL); 2 and 9, Marker V (Boehringer Mannheim) size standard; 3,7 and 8; normal controls; and 4-6, intron 16 C→T variants. C, detection of the intron 17 6 bp deletion. The 417 bp fragment was amplified using ^{32}P-end-labeled primer JS24 and subsequently digested with AluI. The resulting fragments were resolved using a 5% denaturing polyacrylamide gel and detected by autoradiography. Lanes are 1-4, FZG family; 5, HEU family subject; and 6-9, normal controls. D, detection of the intron 16 T→C variant by direct sequencing. E, detection of the codon 708 Gly→Gly variant by direct sequencing. F, detection of APP variants by SSCP. Samples are as follows: 1 and 9, codon 708 variant; 2, codon 717 Val→Ile; 3, normal control; 4, 6 bp deletion variant; 5, codon 693 Gly variant; 6, codon 717 Phe FAD mutation; 7, intron 16 MnlI variant; and 8, intron 16 T→C variant.

Fig 5. Multipoint analysis of linkage of FAD to D21S13 and D21S1/S11 In Volga German families and non-Volga German early-onset kindreds. D21S13 was analyzed as a 4-allele marker and D21S1/S11 as 2 4-allele systems separated by 0% recombination. Non-Volga German early-onset families (solid line) were AM, KG, L, 603, SNW and V and Volga German families (dashed line) were E, H, HB, HD, KS, R and W. The recombination distance used between D12S13 and D21S1/S11 was 0.12. Map distances were calculated using Haldane's formula.

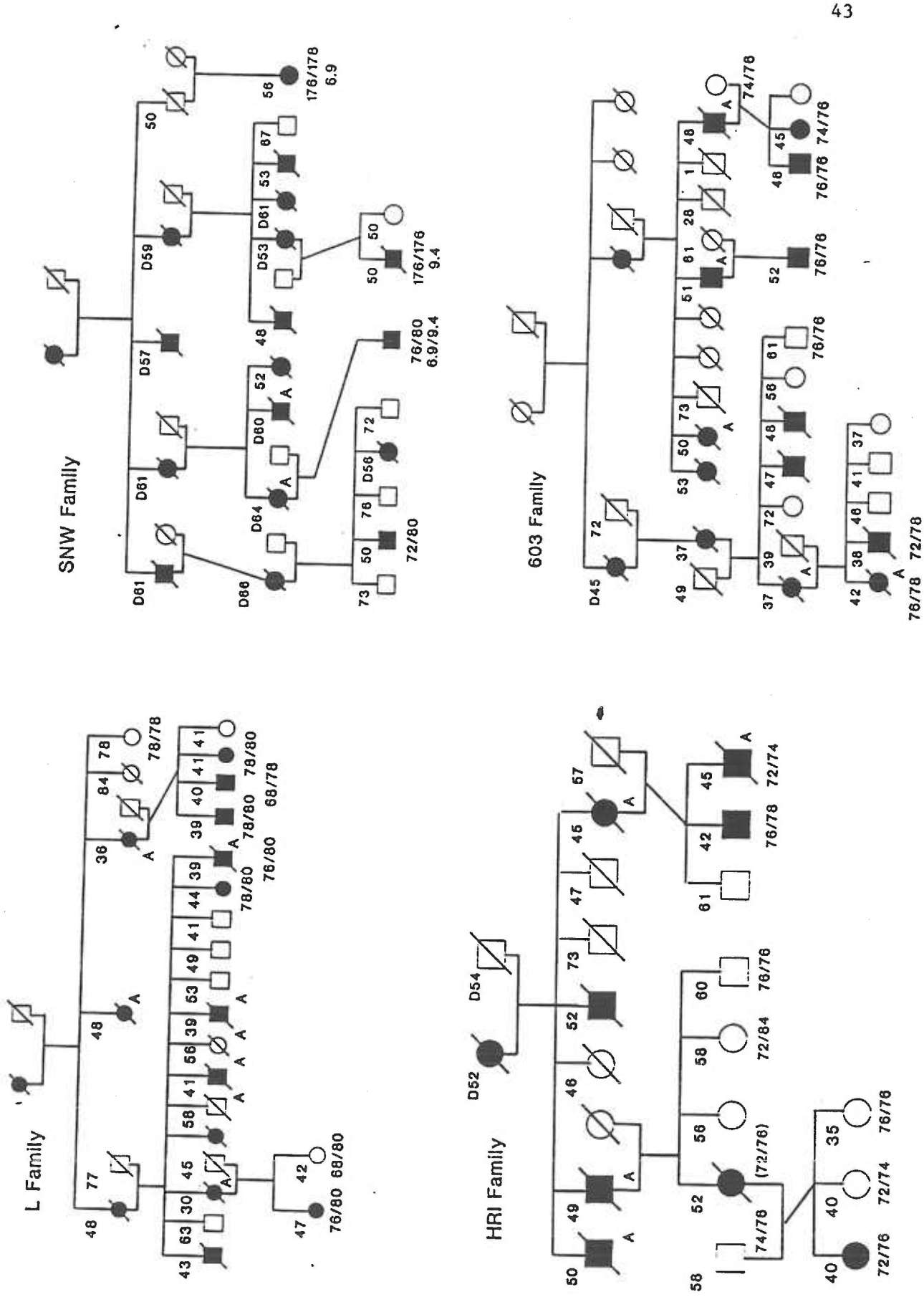
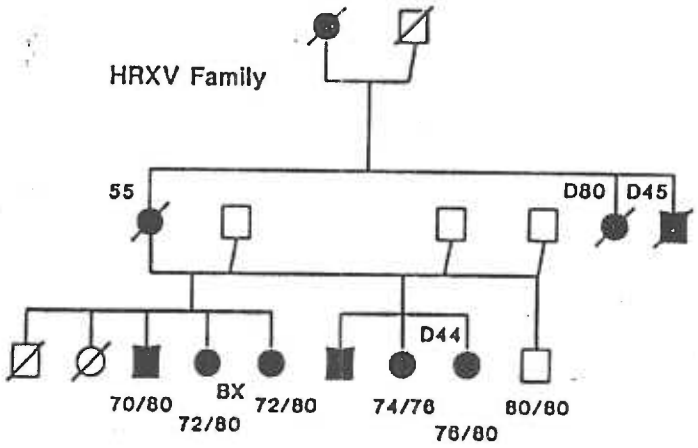
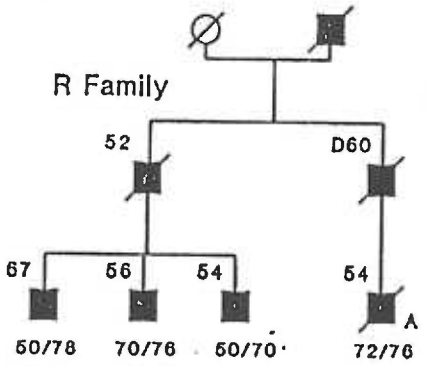
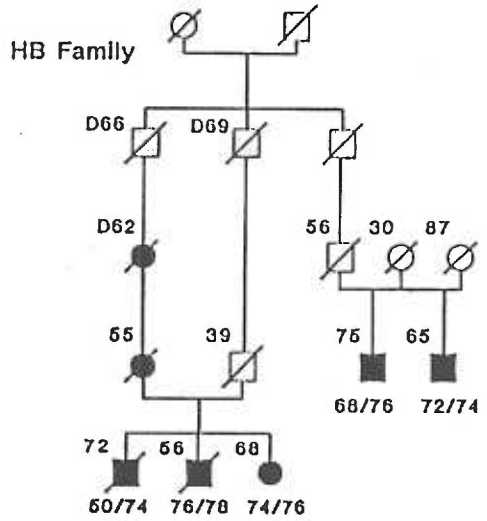
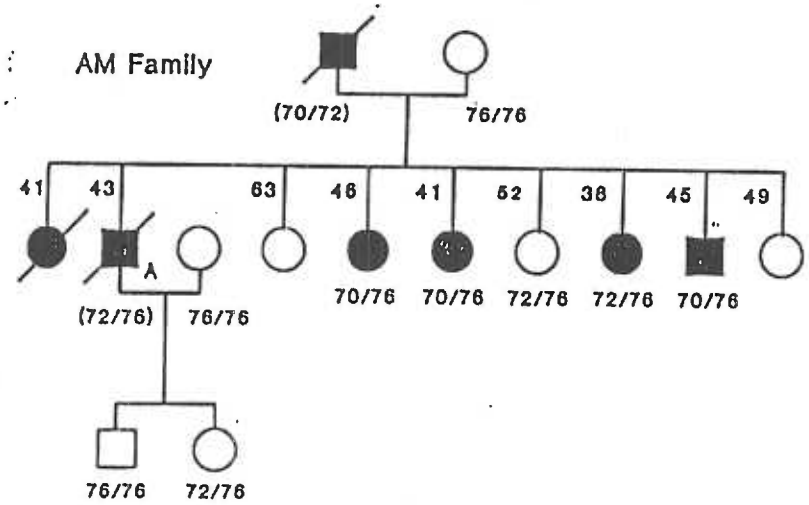


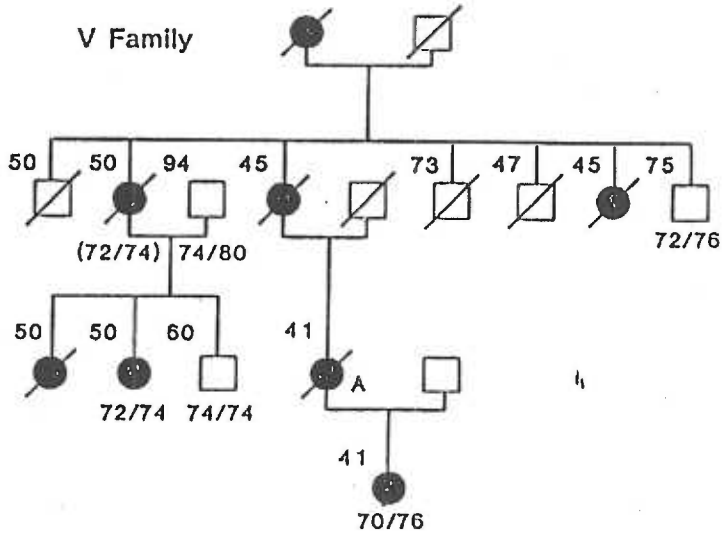
Fig. 1

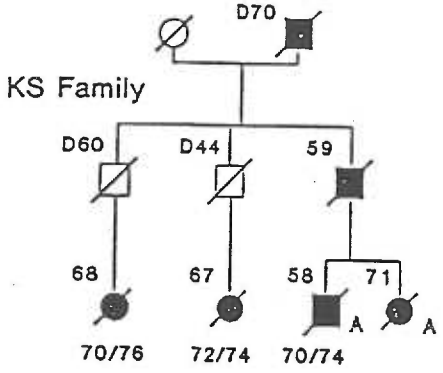

HRXV Family
55
D80
D45
D44
70/80
72/80
BX
72/80
74/76
76/80
80/80


R Family
52
D60
67
56
54
54
A
60/78
70/76
60/70
72/76


HB Family
D66
D69
D62
56
30
87
55
39
75
65
72
56
68
68/76
72/74
60/74
76/78
74/76


AM Family
(70/72)
76/76
41
43
63
46
41
52
38
45
49
A
(72/76)
76/76
70/76
70/76
72/76
72/76
70/76
76/76
72/76


V Family
50
50
94
45
73
47
45
75
(72/74)
74/80
72/76
50
50
60
41
A
72/74
74/74
41
70/76


KS Family
D70
D60
D44
59
68
67
58
71
A
A
70/76
72/74
70/74

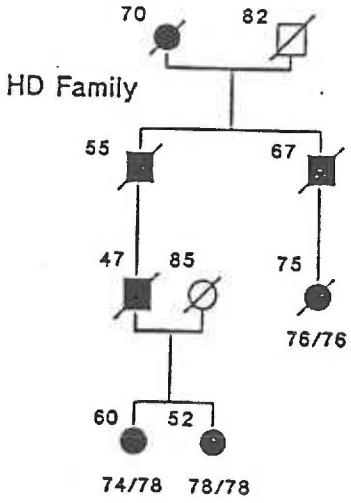
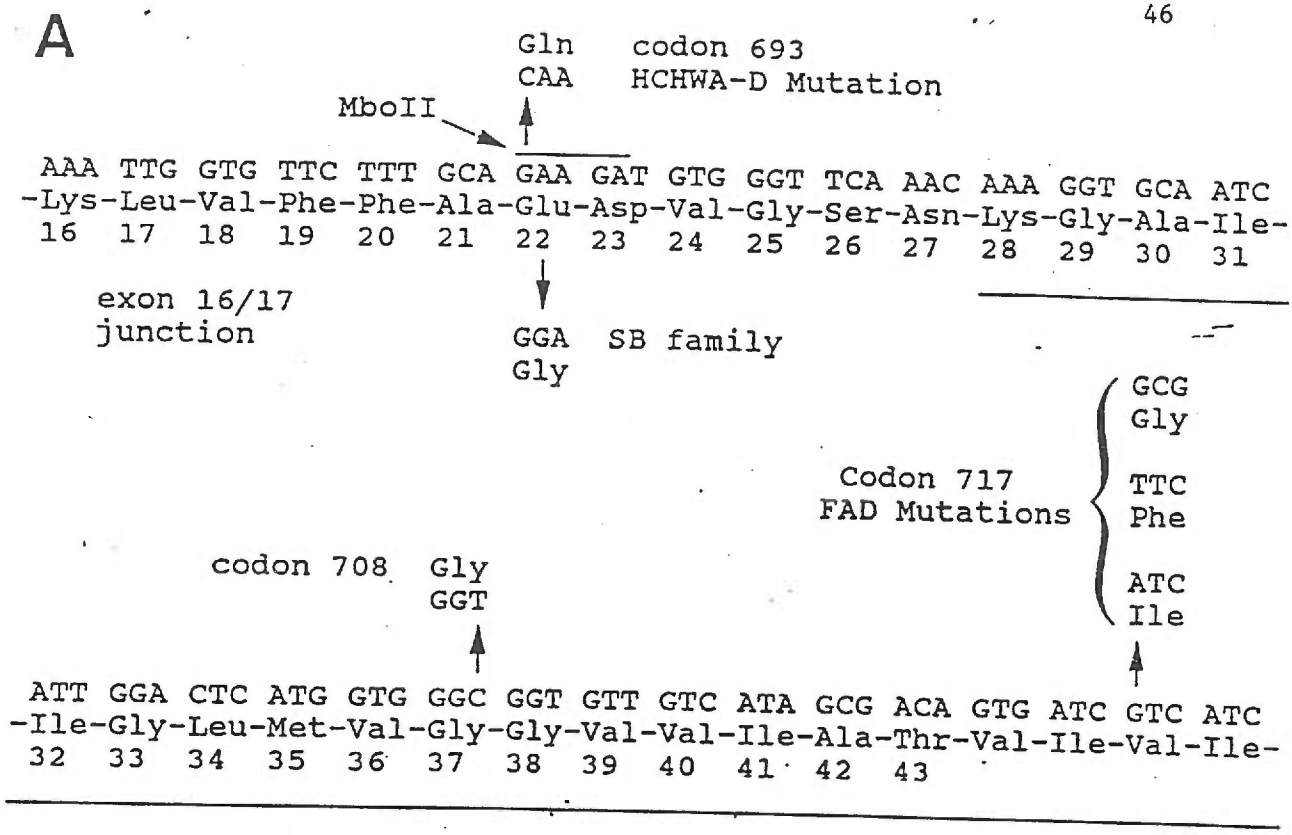
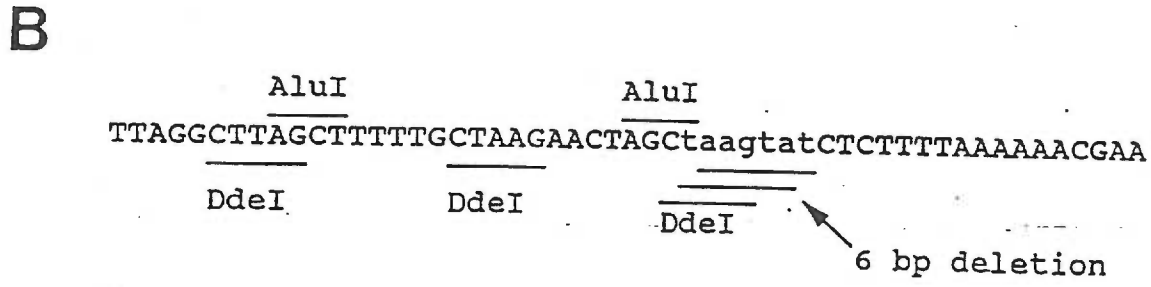
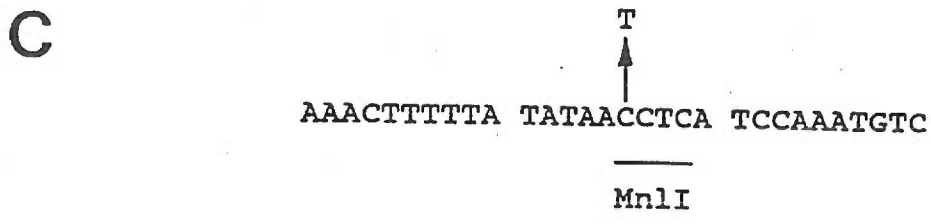

HD Family
70
82
55
67
47
85
75
76/76
60
52
74/78
78/78

FIG 1 (con't)


A
Gln codon 693
CAA HCHWA-D Mutation
MboII
AAA TTG GTG TTC TTT GCA GAA GAT GTG GGT TCA AAC AAA GGT GCA ATC
-Lys-Leu-Val-Phe-Phe-Ala-Glu-Asp-Val-Gly-Ser-Asn-Lys-Gly-Ala-Ile-
16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
exon 16/17 junction
GGA SB family
Gly
GCG
Gly
Codon 717
FAD Mutations
TTC
Phe
ATC
Ile
codon 708 Gly
GGT
ATT GGA CTC ATG GTG GGC GGT GTT GTC ATA GCG ACA GTG ATC GTC ATC
-Ile-Gly-Leu-Met-Val-Gly-Gly-Val-Val-Ile-Ala-Thr-Val-Ile-Val-Ile-
32 33 34 35 36 37 38 39 40 41 42 43


B
AluI
AluI
TTAGGCTTAGCTTTTTGCTAAGAACTAGCtaagtatCTCTTTTAAAAAACGAA
DdeI
DdeI
DdeI
6 bp deletion


C
T
AAACTTTTTA TATAACCTCA TCCAAATGTC
MnlI

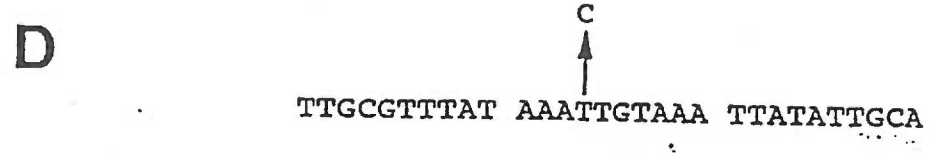

D
C
TTGCGTTTAT AAATTGTAAA TTATATTGCA

FIG 2

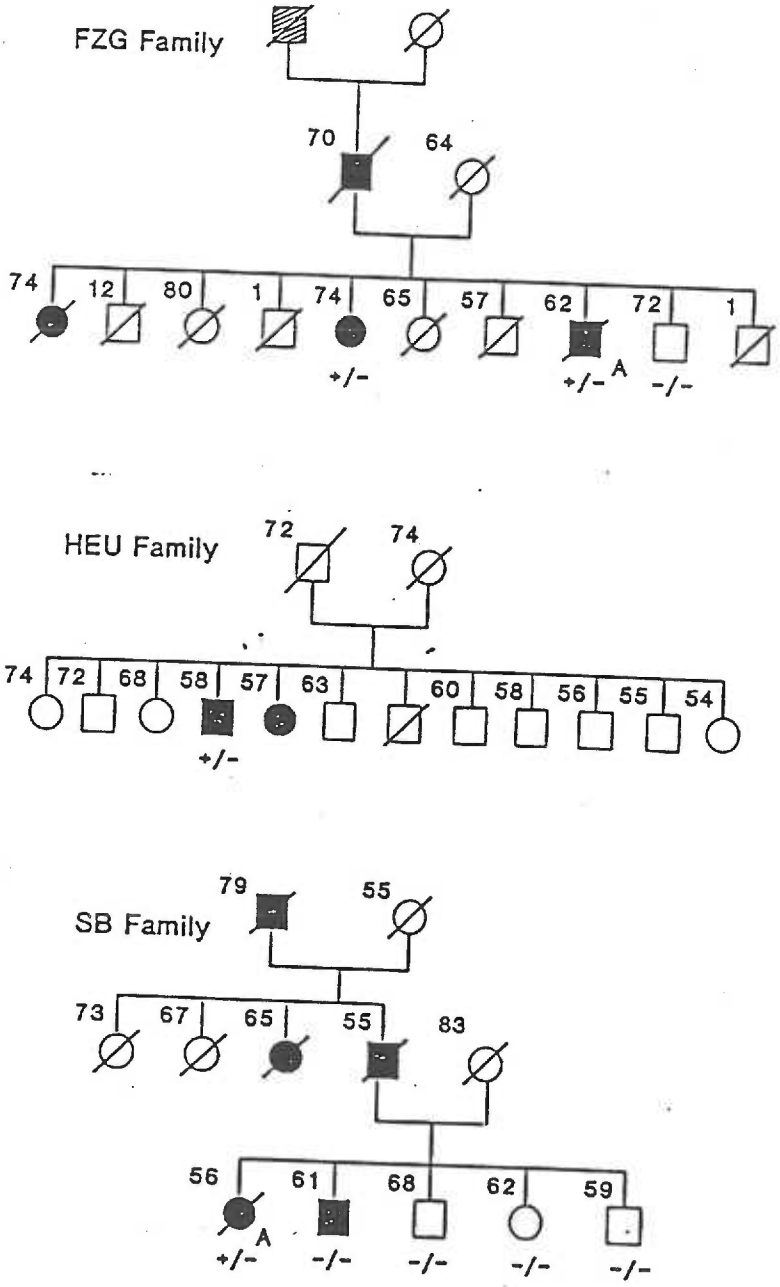

FZG Family
70
64
74
12
80
1
74
65
57
62
72
1
+/-
+/-
A
-/-
HEU Family
72
74
74
72
68
58
57
63
60
58
56
55
54
+/-
SB Family
79
55
73
67
65
55
83
56
61
68
62
59
+/-
A
-/-
-/-
-/-
-/-

FIG 3

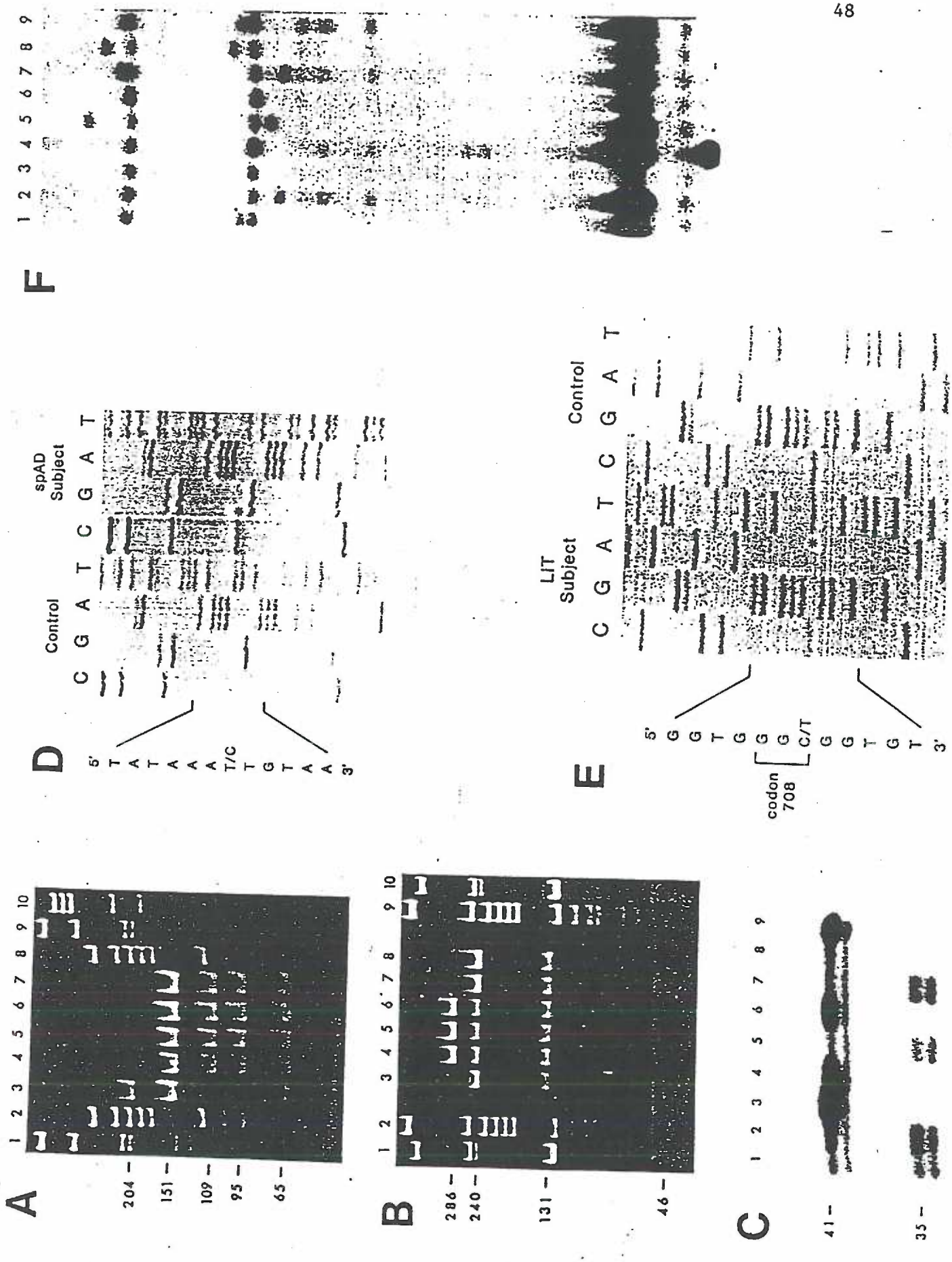

A
1 2 3 4 5 6 7 8 9 10
204 –
151 –
109 –
95 –
65 –
B
1 2 3 4 5 6 7 8 9 10
286 –
240 –
131 –
46 –
C
1 2 3 4 5 6 7 8 9
41 –
35 –
D
Control
spAD Subject
C G A T C G A T
5' T A T A A A T/C T G T A A 3'
E
LIT Subject
Control
C G A T C G A T
5' G G T G G C/T G G T G T 3'
codon 708
F
1 2 3 4 5 6 7 8 9

FIG. 4

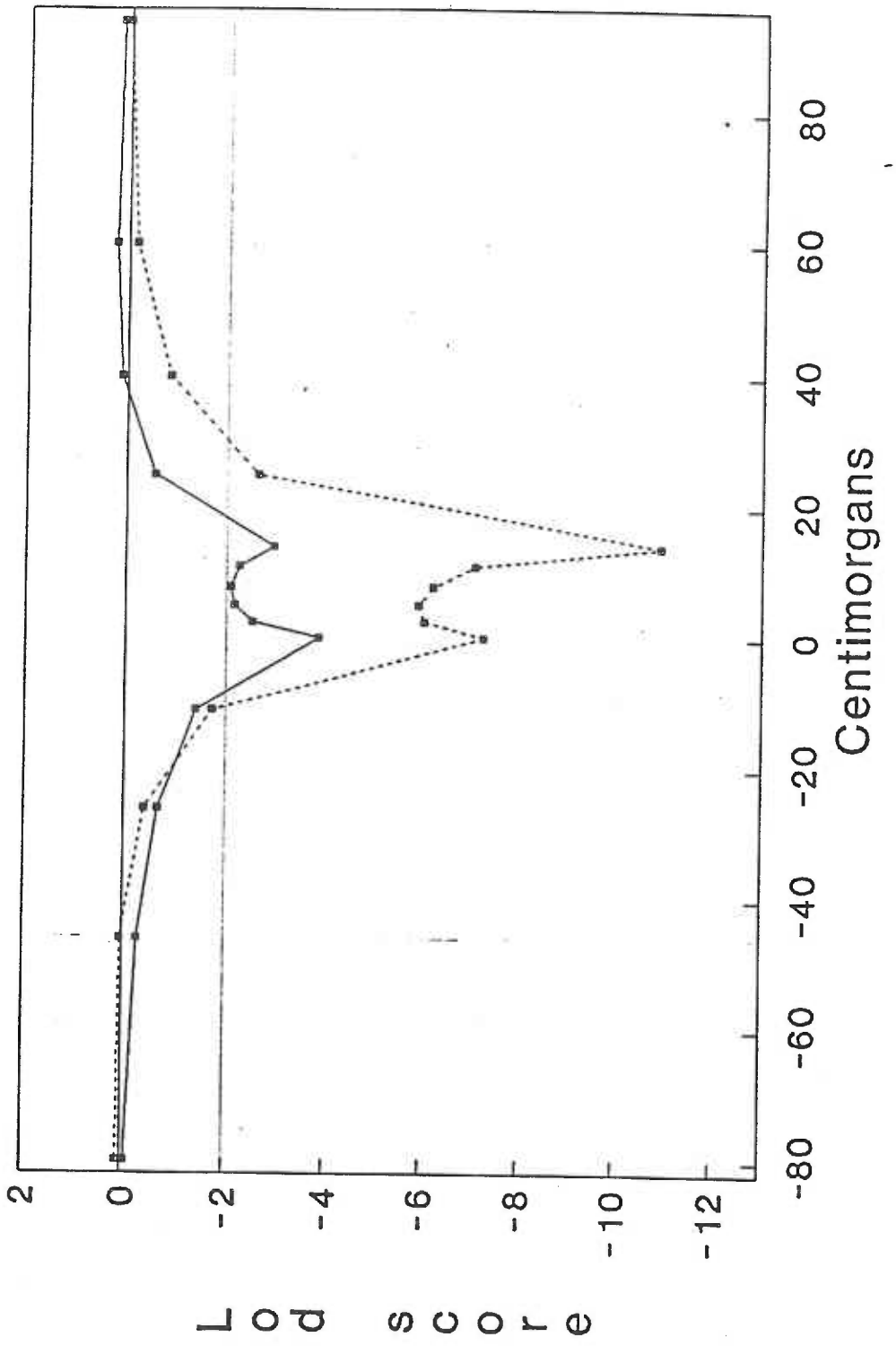


Figure 5