A Genetic Linkage Index Map of the Human Chromosome 11q Composed Mainly of Microsatellite Markers

by

Xue Ya Hauge

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 requirement for the degree of

Doctor of Philosophy

October 1992

APPROVED:

[redacted]

.....
(Professor in Charge of Thesis)

[redacted]

.....
(Chairman, Graduate Council)

TABLE OF CONTENTS

---

* see Appendix

List of Figures

List of Figures cont.

List of Tables

List of Abbreviations

A --- adenine
bp--- base pair
C --- cytosine
CEPH- Centre d'Etude du Polymorphisme Humain
cM--- centiMorgan
cpm-- count per minute
dsDNA double strand DNA
G --- guanine
Het-- heterozygosity
INS-- interspersed sequence
kb--- kilobase pairs
LOD-- log of odd
LSO-- locus specific oligo
M --- mole
Mb--- megabase pairs
mM--- millimole
ml--- milliliter
mg--- milligram
mCi-- millicurie
nt--- nucleotide
PCR-- polymerase chain reaction
PIC-- polymorphic information content
pM--- picomole
RFLP- restriction fragment length polymorphism
rpm-- rotation per minute
SCL-- single colony lysis
SSM-- slipped-strand mispairing
ssDNA single strand DNA
STS-- sequence tagged site
std-- standard
T---- thymine
VNTR- variable number of tandem repeat
ug--- microgram
ul--- microliter
uM--- micromole
uCi-- microcurie

Abstract

The human chromosome 11q contains about 2.7% of the genome and may contain as many as 2,700 genes. Only a very small proportion of those genes (about 1%) have been mapped to the 11q, even fewer have been cloned and sequenced. As a participant in the Human Genome Project, we have taken the responsibility for constructing an index map of chromosome 11q. Here we report isolation, characterization of highly polymorphic markers and construction of the index map of chromosome 11q. Eight highly informative simple sequence repeat polymorphisms have been isolated and mapped to specific bands on the long arm of human chromosome 11 by using somatic cell hybrid panels, in situ hybridization and linkage analysis. Five of them are index quality markers with heterozygosity, or PIC, greater than 0.70. All of them can be genotyped by PCR, and two of them (D11S527/S534) have been typed simultaneously by multiplex PCR. An index map of human chromosome 11q has been constructed mainly based on index markers developed by us and by two other groups. It covers the entire long arm of chromosome 11 except for the very distal region (11q25-qter). The sex-averaged index map of the chromosome 11q spans 82 cM, whereas the sex-specific index maps span 118 cM and 53.4 cM for female and male , respectively. The genetic distance between two adjacent

index markers varies from 1.4 cM to 14.7 cM. The order of the index markers is cen-S228/S436-sMSH3-sHI1-D11S97-INT2-D11S534-D11S533/S527-D11S388-sMSH2-D11S35-Mfd231-DRD 2-D11S144/S351-D11S490-6d11-D11S439-qter. Both possible germline and somatic mutations have been found at two of the index marker loci (D11S534/S439). The Mendelian inheritance of one new mutation was observed. We see no bias towards longer and more polymorphic microsatellite markers. The nature of all mutations is either gain or loss of two nucleotides. Since this feature corresponds very well with one of the features of a slipped-strand mispairing event, we believe that the new mutations described here were generated by the slipped-strand mispairing mechanism. We have described a simple, powerful method which allows us to directly detect a slippage event in simple sequence tandem repeat regions during the polymerase chain reaction. All three major characteristics expected for slippage have been observed in the dinucleotide repeat region examined. The slippage is biased towards dinucleotide rather than the tetranucleotide repeat. We were able to exclude the possible explanations that recombination and/or "out of register" events during PCR would generate the anomalous products. Our conclusion is that the anomalous products are probably generated by slippage. The level of the length polymorphisms between dinucleotide and tetranucleotide repeats are very different. Seventy five

percent of polymorphisms seen at D11S527 which we have examined are due to variations in the dinucleotide region. Only 25% of polymorphisms are due to variations in both dinucleotide and tetranucleotide regions. This observation, again, supports the hypothesis that the slippage may play a major role in the repetitive DNA sequence evolution by producing a larger number of length polymorphisms.

ACKNOWLEDGMENTS

I wish to thank Dr. Michael Litt for his valuable discussion on this manuscript and Drs. Susan B.Olson, David Browne and Ms Marianne Otoshi for their careful reading of the manuscript.

Part I

Introduction

The human genome consists of 3,300 Megabase (Mb) pair of DNA and may encode as many as 100,000 genes. Of them, only about 4,600 (<5%) are represented in Mendelian Inheritance in Man (Mckusick 1990); fewer (1.5%) have been mapped to specific chromosomes or subchromosomal regions; even fewer (0.6%) have been cloned and sequenced.

The traditional approach to the study of inherited disorders and ultimately to cloning disease-related genes has started, in most cases, with the identification and characterization of specific proteins and their abnormal counterparts. This approach has been fairly successful in the early stage of searching for the disease-related gene in man. However, since the biochemical defects or protein products of many inherited disorders are poorly understood, and corresponding animal models of the diseases are lacking, the majority of Mendelian disorders have to be dissected by using a completely different approach—the "positional cloning" method. Positional cloning involves 1) utilizing a well constructed genetic linkage map of human genome to detect the linkage of polymorphic DNA markers and an inherited disorder segregating in affected families; 2) defining the precise

position of a gene on a particular chromosome or a subchromosomal region, and 3) identifying a specific gene within this region in which the mutations are correlated with the disease. It is clear that the construction of a dense genetic linkage map with a large number of restriction fragment length polymorphism (RFLP) and variable number of tandem repeat (VNTR) markers, especially microsatellite markers, is essential to accomplish of the enormous undertaking of identification and characterization of disease related genes.

The RFLP marker detects the variation of a particular kind of DNA sequence called the restriction enzyme recognition sites in the human genomic DNA. It generally exists in only two forms—presence or absence of the site. This implies that, at most, about 50% of individuals will be heterozygotes at the DNA marker locus (Botstein et al. 1980). The VNTR marker, on the other hand, detects the variation of the length of tandem repeats which have an unknown function and are hypervariable among individuals in the population (Jefferys et al. 1985a, 1985b, 1985c). Likewise, a microsatellite marker detects the length variation of a tandem repeat, which is considerably shorter than a VNTR and will be discussed in more detail later.

The usefulness of any marker depends largely on how many alleles (or variants) it displays in the population. the more alleles it has, the more likely the individual

tested would carry two different alleles (heterozygote) at this marker locus. In the two-locus or multilocus linkage analysis, if the individual tested carries two different alleles at all of these marker loci, we will be able to determine the "phase": how the alleles are distributed between the two homologous. Consequently, we can detect unambiguously the recombination between markers in his (or her) offspring. In addition to the heterozygosities of the markers, the three generation families with large pedigrees are ideal for the linkage analysis. This is because the presence of many children means that the parents' chromosomes can be followed through a large number of meioses, generating more accurate estimates of recombination frequency than could be had from families with few children. The "phase" of the parent can be usually inferred from that of the grandparents. Therefore, CEPH (Centre d'Etude du Polymorphisme Humain) has collected DNA samples from 60 extensive pedigrees and provided these samples to the investigators around the world.

In the past nine years, more than a dozen important inherited disorders have been mapped by utilizing linkage analysis plus cytogenetic analysis. About half of these have been cloned and sequenced by using the "positional cloning" strategy, including familial polyposis coli on chromosome 5 (5q21-q22) (Kinzler et al. 1990), cystic fibrosis on chromosome 7 (q31.3-q32) (Rommens et al.

1989), retinablastoma on chromosome 13 (q14.1-q14.2) (Friend et al. 1986), neurofibromatosis type 1 on chromosome 17 (q11.2) (Wallace et al. 1990), myotonic dystrophy on chromosome 19 (q13.2-q13.3) (Harley et al. 1992), one form of Alzheimer's disease on chromosome 21 (q21.3-q22.05) (Goate et al. 1991), Fragile X syndrome (q27.3) (Oberle et al. 1991), Duchenne muscular dystrophy (p21.2) (Kunkel et al.) and chronic granulomatous disease (p21.1) (Royer-Pokora et al. 1986) on the X chromosome.

Linkage has also been used to localize a number of human genetic diseases to the long arm of chromosome 11 (11q), including multiple endocrine neoplasia type 1 (Larsson et al. 1988), tuberous sclerosis (Smith et al. 1990), ataxia telangiectasia (Gatti et al. 1988; Ziv et al. 1991) and atopy (asthma and rhinitis) (Cookson et al. 1989). An extensive linkage analysis also made it possible to exclude the presence of a susceptibility gene for insulin-dependent diabetes mellitus from chromosome 11q (Hyer et al. 1991). At 11q21-q22 region, breakpoints found in balanced translocations, t(1;11) and t(9;11), associated with schizophrenia and bipolar manic depression, have been defined by polymorphic DNA markers in that region (St. Clair et al. 1990). Further toward the telomere, the 11q22-q23 region contains at least 16 functional genes, several disease loci and clusters of break points involved in a number of leukemias and malignant tumors, including acute intermittent porphyria

(Namba et al. 1991), tuberous sclerosis, Jacobson syndrome (Fryns et al. 1986), apolipoprotein complex A1-C3-A4 (McKusick 1987), Ewing sarcoma (Selleri et al. 1991), peripheral neuroepithelioma (Budarf et al. 1989), Dopamine D2 receptor (Grandy et al. 1989b), progesterone receptor (Roussean-Merck et al. 1987), ataxia telangiectasia complementation group A and group C, several members of the immunoglobulin super family (Rowley et al. 199), acute lymphoblastic leukemia (ALL), acute monoblastic leukemia (AML) (Rowley et al. 1990; Tunnacliffe et al. 1990), N-CAM (Nguyen et al. 1985) and two prooncogenes Ets (de Taisue et al. 1984) and Cbl-2 (Wei et al. 1990) (Fig.I-1). Since the human genome contains about 100,000 genes and chromosome 11 consists of 4.51 per cent of the genome, the estimated number of genes on the entire chromosome 11 and 11q are about 4500 and 2700, respectively.

The chromosomal locations of genes or disease related genes and translocation breakpoints reported so far are strongly biased toward 2 major Giemsa (G-) negative bands, namely 11q13 and 11q23. It has been well documented that most genes characterized are clustered in the G negative-R positive bands in the human genome. Furthermore, the human genome compositional map has shown some very similar and interesting data, where the concentration of isochore family H3 was used to detect the distribution pattern of genes in the human genome

(Saccone et al. 1992). It has been demonstrated that isochore family H3 has the highest concentration of genes and of CpG islands. The cytogenetic location of isochore family H3 is mainly at telomere T-bands and G negative bands. Two regions of human chromosome 11, 11p15 and 11q13, showed the strongest signals when hybridized with biotinylated family H3 DNA probe. Less intense signals were also found at 11q23 and 11q24-25 regions.

A number of maps for chromosome 11q have been constructed, including genetic linkage maps and cytogenetic maps. A preliminary linkage map of chromosome 11 was reported by Leppert et al. (1987). It was followed by a more detailed linkage map of the chromosome 11q in which 31 framework loci were included which covered the region from 11q13 to 11q24-25 (Julier et al. 1990). However, since most of those framework marker loci are RFLPs, their heterozygosity is less than 0.60, with two exceptions (one is a haplotyped marker and another is a VNTR marker). Additionally, they are clustered in two chromosome bands, 11q13 and 11q23, and the odds against inversion of orders of 10 loci are less than 100:1, making linkage analysis in certain regions more difficult. Therefore, it is quite clear that the current linkage map has two major drawbacks. First, most of the preexisting markers have low heterozygosities or polymorphism information content (PIC). The markers which detect the restriction fragment length polymorphism

of the genomic DNA are usually not highly polymorphic, given that those markers only have two alleles at each locus, and, therefore the maximum value of PIC for the locus can barely reach .35. On the other hand, the markers which detect variable numbers of tandem repeats (VNTR) of human genomic DNA are usually highly polymorphic, but they are predominantly located near the telomere region of chromosomes (Jefferys et al. 1985a, 1985b & 1985c; Nakamura et al. 1987). Furthermore, the alleles of VNTR markers tend to have complex banding pattern, potentially complicating the genotyping process. Second, the preexisting markers are not evenly distributed along the long arm of chromosome 11, leaving several gap regions on the linkage map. Those gap regions are centromere region, 11q14, 11q21-22, 11q24-25 and telomere region.

Two groups have constructed cytogenetic maps and/or physical maps of the long arm of chromosome 11 where about 228 chromosome 11q specific cosmids including 31 NotI cosmid linking clones are localized to defined regions, either by using somatic hybrid panels or by high resolution <u>in situ</u> hybridizations (Lichter et al. 1990; Tokino et al. 1991; Hori, T.-A. et al. 1992; Iizuka et al. 1992; Hermanson et al. 1992). Again, the major drawback to the current cytogenetic and/or physical maps is that most cosmids are densely concentrated in 11q13 and 11q23 regions, the same situation that we have seen

in the linkage map. Additionally, some of those cosmids which have been detected to be polymorphic show low heterozygosities. Nonetheless, those cytogenetic and/or physical maps provide excellent references for us to construct a linkage index map and, eventually, a high resolution linkage map of the long arm of human chromosome 11. They also provide us the reagent for isolation of a new type of highly polymorphic DNA markers in the gap regions on the index map.

In addition to traditional RFLP and VNTR markers, a recent report by Iizuka et al. described another type of DNA polymorphism—the Alu repeat polymorphism, which can be detected by Single Strand Conformation Polymorphism (SSCP) analysis (Orita et al. 1989). They have isolated 7 Alu repeat polymorphic markers from human chromosome 11 with low PICs.

One of the short term goals of the Human Genome Project is to establish the genetic index map for each human chromosome. An index map consists of a number of index loci, which should have heterozygosity of 0.7 or higher, and should be evenly distributed along the arms of chromosomes. The distance between two adjacent index loci should be no greater than 10 to 15 centiMorgan (Beats et al. 1991). The establishment of a dense linkage map, particularly with evenly distributed index markers, of chromosome 11q would greatly benefit all investigators in the human genetic community. With such a dense

linkage map, the chromosome location of newly discovered linkage could be confirmed or excluded easily; the complex, heterogenous disease and rare recessive disease would be analyzed much more efficiently because as the informativeness of the markers increases, the number of the meioses of individuals from the affected families required will dramatically decrease; several nearby starting points would be available for efforts to clone the disease gene; prenatal or presymptomatic diagnosis of individuals at risk would become more accurate; identification of translocation breakpoints, study of gene conversion and recombination event would be possible (Donis-Keller et al. 1987). It would be impossible to construct an index map consisting of traditional RFLP and VNTR markers for the reasons mentioned above. Therefore, it is clear that new types of DNA polymorphisms have to be explored and developed to meet the challenge.

The discovery of polymorphic simple sequence repeats in the human genome has revolutionized genetic linkage studies (Weber and May 1989; Litt and Luty 1989). Simple sequence repeats are usually di-, tri- and tetranucleotide repeats with highly varied numbers of repeat units. Among the different types of dinucleotide repeats, the poly CA repeat is the most abundant. It has been estimated that there are as many as 10^5 (n>10) copies of them in the human genome, and the average distance between two poly CA tracts is about 30 kb

(kilobase pairs). Additionally, poly CA repeats are randomly distributed in the genome and located mainly in the non-coding region of the genome or within the introns of genes. The function of the dinucleotide repeat is unknown, although there are some indications that they stimulate DNA recombinations (Slightom et al. 1987). The poly CA repeat markers have been called microsatellite repeat markers because its repeat unit contains only two nucleotides. The variation in the number of repeat units in a particular poly CA tract, or polymorphism, is transmitted through generations in Mendelian codominant fashion.

For the purpose of investigating their informativeness, microsatellite repeats (CA/GT repeat in particular) can be grouped into four classes: 1) perfect repeat: It has no interruption in a run of repeats; 2) interrupted repeats: two or more short stretches of repeats in close vicinity and spaced by non-repeat sequences; 3) compound perfect repeat: more than one type of microsatellite repeat juxtaposed to one another. These microsatellite repeats are different either in nucleotide composition, such as purine-purine, pyrimidine-pyrimidine, purine-pyrimidine, or in the number of nucleotides in one repeat unit, such as tetra-, tri- and dinucleotide repeats (Weber 1990); 4) compound interrupted repeat: the criteria for the compound interrupted repeat is basically the same as that for the

compound perfect repeat, except that they are spaced by non-repeat sequences. Weber (1990) has proposed a similar grouping method.

In general, a positive correlation between the degree of polymorphism and the length of dinucleotide repeat can be found in the marker that shows polymorphism. The most striking and important feature of microsatellite DNA markers is high informativeness. The microsatellite DNA markers have, on the average, PIC values of 0.6, twice as high as those of RFLPs. According to our own data, 75 per cent of the markers we have described to date have PIC values of 0.6 or higher. The maximum PIC value for microsatellite markers to date is over 0.90. In addition, there are several other advantages in using microsatellite DNA markers. First, much less time and reagents for genotyping a lager number of samples, e.g. 40 CEPH (Centre d' Etude du Polymorphisme Humain) reference families, are required. Since each locus can be genotyped by using polymerase chain reaction (PCR), it has the potential of being amplified simultaneously with other loci by multiplex PCR method. Those markers also have the potential of being semi-automated. Second, much less DNA sample from each individual is needed for genotyping because the PCR can amplify a single copy of DNA molecule to a million-fold, facilitating the linkage analysis in affected families with rare diseases. Finally, since the power of

resolution of the denaturing gel is so high that one nucleotide difference can be detected, it leaves much lees doubt in the allele assessment when compared with VNTR markers where DNA fragments with a 2% size range may appear to comigrate and be assessed as the same allele.

Microsatellite DNA markers are also qualified Sequence-Tagged Site (STS) markers, serving as landmarks to define the position of any cloned DNA segment on the physical map (Olson et al. 1989). The sequence-tagged sites are short tracts of single-copy DNA sequences that can be recovered at any time by polymerase chain reaction. By using the STS markers, the construction of a physical map of the long arm of human chromosome 11 would be like determination of the order and spacing of DNA segments; each of them has an uniquely defined STS. This will help to merge mapping data gathered by diverse methods (including physical, genetic, and cytogenetic <u>in situ</u> hybridization methods) from different laboratories into a consensus physical map.

Since microsatellite DNA markers are abundant in the human genome, highly informative and easy to be genotyped through the random individual in the population or members of the affected families, they have become powerful tools in linkage studies and are the most promising candidates for index markers. Ever since the first two papers on poly CA repeat polymorphisms in humans were published (Weber and May 1989; Litt and Luty

1989), hundreds, maybe thousands, of microsatellite markers have been developed. They have facilitated the localization of a number of disease-related genes, including Friedreich ataxia (9q13-q21.1) (Fujita et al. 1990), facioscapulohumera muscular dystrophy (FSHD) (4q35-qter) (Wijmenga et al. 1990), an X-linked mental retardation disorder (Xq26) (Huang et al. 1991) and limb-girdle muscular dystrophy (5q22.2-q31.3) (Speer et al. 1992). Linkage maps that consist only of microsatellite markers have also been constructed for particular chromosomes, including chromosome 14 and 20 (Wang and Weber 1992; Hazan et al. 1992). In addition to localization of disease-related genes and construction of linkage maps, carrier detection, prenatal diagnosis and family counseling have been improved and simplified by the development of highly polymorphic microsatellite markers in several well characterized disease-related genes, including cystic fibrosis (Zielenski et al. 1991), Duchenne and Becker muscular dystrophies (Clemens et al. 1991). It is undoubtedly true that application of the microsatellite markers will speed up the completion of the Human Genome Project.

The telomere is of interest to us for several reasons. First, there is no polymorphic DNA marker developed in the telomeric region of chromosome 11. Since human telomeres define the chromosome ends, without telomeric markers, the linkage map of chromosome 11 would

be incomplete. Second, it has been reported that there are genes found close to the telomere in Drosophila melanogaster and Saccharamyces cerevisiae (Lufzel-Schwab et. al. 1986 & Charron et. al. 1988). In humans, one of two mutually exclusive locations for Huntington's disease gene is close to the telomere of the chromosome 4 shortarm (p) (Pritchard, C. et al. 1990; Bates, G.P. et al. 1991), and polycystic kidney disease gene and alpha-globin gene are located near the telomere of chromosome 16p (Reeders et al. 1988; Wilkie et al. 1991). Therefore, polymorphic DNA markers isolated from the telomere of chromosome 11 would be useful for localizing a disease gene by linkage analysis, if such a gene exists in the chromosome 11 telomeric region. Third, much of our knowledge of the recombination events in the telomeric region of humans comes from the studies of pseudoautosomal regions, which are located at the very ends of sex chromosomes; the corresponding data of autosomal telomere are lacking. In the pseudoautosomal region of the Y chromosome, there is no linkage between the telomeric locus (DXYS14) and sex the phenotype observed (Weissenbach et al. 1987; Rouyer et al. 1986). This means that there is 50% recombination between the telomere region and the putative testes-determining region close to it. The human sex-determining region is about 2.7 Mb away from the telomere of the Y chromosome. Therefore, the recombination frequency between these two

closely juxtaposed regions is about 18.5 times higher than that of the average chromosomal region with the same physical distance. On the other hand, in the chromosome 4p16 region, the recombination rate observed per unit of physical distance decreases dramatically as one moves closer to the telomere. There has been no gross excess of recombination expected for a telomeric region found in 4p16 region (Bates et al. 1991).

As a participant in the Human Genome Project, we have taken the responsibility for constructing an index map of human chromosome 11. To this end, the work reported herein mainly involves development of highly polymorphic microsatellite markers on the long arm of human chromosome 11 and the use of these markers to construct an index map. The current index map of chromosome 11q consists of a few preexisting RFLPs which have been haplotyped, or VNTR markers. There are several gaps on this map, including the centromere region, 11q14, 11q21-q23, 11q24-q25 and the telomere region. Those gaps have to be filled with highly polymorphic markers. Therefore, the folowing goals were for this projet:

1. Isolation and characterization of highly polymorphic microsatellite (di-, tri- and tetranucleotide repeat) DNA markers. These markers should have heterozygosity of .70 or higher and should be located in the gap regions of the current chromosome 11q index map

integrating them into the preexisting chromosome 11q index map.

2. Genotyping of highly polymorphic markers, developed by us, for all CEPH reference families by utilizing multiplex PCR and multiplex sample loading methods; generating genotype/haplotype data for each CEPH individual on the DNA marker loci.
3. Providing of further evidence for the mechanism giving rise to "spurious bands", which have been encountered by many investigators and still remains as a mysterious phenomenon.
4. Providing of more accurate data on spontaneous mutations of highly polymorphic microsatellite markers.
5. Isolation of polymorphic DNA markers in the telomere/protelomere regions of the long arm of human chromosome 11.

Section I

Isolation and Charaterization of Highly Polymorphic Microsatellite Markers from the Long Arm of Human Chromosome 11

Materials and Methods*

1. Cell Lines and Cosmid Clones

Somatic cell hybrid lines MC-1,J1-44, J1-11 and TGD5D1-1 have been described previously (Maslen et al. 1988; Jones et al.1983; Jones et al. 1984) (Fig I-3 Cosmid clones mapped to the long arm of chromosome 11 were from two independent cosmid libraries. One set of was isolated from a genomic library constructed in cosmid vector sCos-1 (Evans and Lewis,1989) by using DNA from a somatic cell hybrid TG5D1-1 carrying 11q13-11qter as the only human material in a mouse erythroleukemia (MEL) cell background; another set of cosmids (cCI11- series) was isolated from a library constructed in cosmid vector pWEX15 by using the J1 cell line, a Chinese hamster x human somatic cell line, which contains the intact human chromosome 11 as the only human material (Tokino et al 1991).

2. Cosmid or Plasmid DNA Preparation

*Experimental methods and solutions with no specified sources in the thesis were referred to the book by Sambrook et al. (1989).

Cosmid and plasmid DNAs were prepared by using the midiprep method developed by Nakamura (Nakamura, personal communication). In summary, 80 ml bacterial culture was grown overnight at 37 °C for 16 hr. The cells were harvested by centrifugation. The pellet was washed with 0.14M NaCl, 10 mM TrisCl⁻ (pH7.5 Cells were then treated with hypertonic buffer which contains 25% (w/v) sucrose, 50 mM TrisCl⁻ (pH 8.0) and 1 mM EDTA (pH 8.0) prior to addition of 20 μl RNase A (10 mg/ml) and 400 μl lysozyme (10 mg/ml). Lysis of cells was performed by using 6 ml lysis buffer containing 0.1% (v/v) Triton X-100, 62.5 mM EDTA (pH8.0) and 50 mM TrisCl⁻ (pH 8.0). The cosmid or plasmid DNA was recovered by addition of a PEG/NaCl solution containing 20% (w/v) polyethylene glycol 600, 20% (v/v) 5M NaCl and centrifugation at 10 K rpm and separated from bacterial genomic DNA by CsCl/EtBr gradient centrifugation at 54 Krpm. Finally, the DNA was recovered by using 0.15M NaAc/95% ethanol, extracted once with equal volumes of phenol/chloroform/isoamy alcohol and once with equal volumes of chloroform/isoamy alcohol and recovered with 0.15M NaAc/ 95% ethanol and then 70% ethanol.

3. Southern Blotting

The DNA was digested with appropriate restriction enzymes and electrophoresis of the DNA samples was

performed in an agarose gel using a sample buffer containing 0.25% (w/v) xylene cyanol and 0.25% (w/v) bromophenol blue. After electrophoresis, the DNA was transferred onto Hybond N^+ nylon membrane by using alkali blotting method according to the instructions of manufacturer (Amershan International plc. Amershan UK). The probe was labeled either by the 5'-end labeling method or by the random labeling method to the specific activity of $1x10^6$ to $1x10^7$ cpm/μg DNA (Feinberg and Vogelstein 1983). The membrane was hybridized with radioactive labeled probe at 65° C overnight. In a hybridization solution of 7% PEG, 10% SDS, 50 mM PO4 and 250 μg/ml sonicated denatured human DNA. The membrane was washed once with 2xSSC, 0.1% SDS at room temperature for 15 min, once with 0.1xSSC, 0.1% SDS at room temperature for 15 min and finally with 0.1xSSC, 0.1% SDS at 60c for 15 min. The membrane was exposed to film for overnight.

4. Subcloning

a. Preparation of the Vector DNA

Plasmid vector pTZ18u was digested with restriction enzyme SalI according to the instructions of manufacturer (BRL). Fifteen μl of 50 ng/μl linearized vector DNA was partially filled-in with 0.4 μl of 4 mM DTT, 0.8 μl of 1 mM dCTP+dTTP, 0.6 unit of Klenow

fragment of polymerase I; incubated at room temperature for 30 min and heated at 65° C for 10 min.

b. Preparation of the Insert DNA

Cosmid or phage to be subcloned was digested with restriction enzyme Sau3A according to the instructions of manufacturer (BRL). Fifteen μl (750 ng) of insert DNA was partially filled-in by adding 0.4 mM DTT, 0.8 μl of 1 mM dGTP+dATP, 0.6 unit of Klenow fragment of polymerase I, incubated at room temperature for 30 min and heated at 65° C for 10 min.

c. Ligation Reaction

In a final volume of 10 μl, approximately 50 ng partial filled-in vector and 2 ng partial filled-in insert DNA were mixed with 2 μl 5x ligase buffer (BRL), 0.5 unit T4 ligase (BRL The mixture was incubated at 16° C overnight. The reaction was stopped by adding 0.5 μl 0.5M EDTA (pH 8.0).

5. Transformation

The procedures of preparation of competent cells and transformation were performed by following the protocols of Hanahan (Hanahan 1985).

6. Colony Screening

The cultured competent cells were plated on the LB agar plate containing 100 μg/ml ampicillin. Colonies

were lifted up by nitrocellulose filters and bacterial DNA was denatured on these filters by incubation in alkali. After a brief washing with neutral buffers, the DNA was bound to the filter by heating in the vacuum oven at 80° C for 1 hr. The filters were then incubated with a (dA-dC)10 oligonucleotide probe which was 5'-end labeled with [gamma-32p]ATP and polynucleotide kinase (New England Biolab) to specific activity of 2.5x108 cpm/μg DNA. Unincorporated [gamma-32p]ATP was separated from the oligomer by passage over a NENSORB^{Tm} column(New England Nuclear Hybridizations and prehybridizations were performed at 60 °C in 6xSSC, 0.1% SDS. Filters were hybridized with probe for 1 hr, washed twice in 6xSSC, 0.1%SDS at room temperature for 15 min and then at 65° C for about 10 min. Autoradiography was performed at -70° C overnight.

7. Plasmid DNA Preparation and DNA Sequencing

After colony screening, positive colonies were picked and grown in 1-5 ml Super Broth containing 100 mg/ml Ampicillin overnight. The plasmid DNA was purified by passing through the Qiagen column according to the instructions of the manufacturer (QIAGEN). The DNA sequencing was performed by using a sequencing kit (United States Biolab) according to the instructions of manufacturer. In general, universal primer and reversed primer provided by the manufacturer were used for

sequencing. In some instances where the size of the insert was too large to obtain the sequence data, so a unique primer was designed to continue to read the sequence.

8. PCR/Multiplex PCR and Detection of PCR Products

PCR primers were designed according to the DNA sequencing data. There are three major criteria in designing PCR primers. First, a set of PCR primers should not have complementary sequences, especially the last two nucleotides at the 3'-end of the primers. Second, the sequence of the primer should have approximately 50% GC content. Third, the primer sequence should be in close vicinity of the target repeat block.

Oligo primers for PCR were synthesized on a 391 PCR-MATE™ DNA synthesizer (Applied Biosystems), deblocked by incubation with concentrated $(NH)_4OH$ for at least 16 hr and lyopholized in Savant Speed Vac. The primers were resuspended in TE and extracted with phenol/chloroform, chloroform and precipitated with ethanol.

PCR were carried out by using either MJ Research® or Perkin-Elmer/Cetus 9600® thermal cyclers and conditions were optimized by a series of experiments. In general, the PCRs were performed in a total volume of 12.5 μl containing 25 ng genomic DNA, 2.5 pmole of each primer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 50 mM KCl, 5 mM $TrisCl^-$, pH 8.3, 0.3 units of Taq polymerase (Perkin-Elmer/Cetus) and

0.01% gelatin. Three parameters for amplification, i.e. the temperature for denaturation, the extension time and the number of cycles, were relatively consistent among different sets of primers. However, the annealing temperature had to be high enough to eliminate non-specific hybridization of primers to genomic DNA templates. Based on the calculated Tm (melting temperature), the annealing temperature was quite different from primer to primer. There have been an increasing number of reports which use organic compounds such as formamide, DMSO and spermidine to lower the Tm or overcome uneven heating of the thermal cycler (Hang et al. 1990; Sarkar et al. 1990). We have used these methods selectively to optimize individual PCR reactions. To visualize the PCR product on the denaturing gel, one of primers was 5'-end labeled with [gamma-32p]-ATP prior to the PCR or [alpha-32p]-dCTP was added to the reaction and then incorporated into newly synthesized DNA strands.

After the PCR amplification, the products were first analyzed on ethidium bromide stained 1.5-2% agarose mini-gel to insure that efficiency and specificity of amplification of the target sequence had been obtained. Then 1.5 excess of sample loading buffer, containing 95% (v/v) formamide, 10 mM EDTA (pH 8.0), 0.1% (w/v) bromophenol blue, 0.1% (w/v) xylene cyano, was added to the PCR products. The mixtures were then resolved on 0.4 mm, 6% polyacrylamide denaturing gels, or in some

instances on 6% denaturing gels containing 32% formamide plus 5.6 M urea. A sequence ladder of a known sequence served as a size standard. Gels were lifted from the glass plate onto Whatman 3MM paper, covered with Saran wrap and dried under vacuum at 80° C. In some cases, the PCR products were transferred onto Hybond N+ membrane by capillary blotting, denatured in the vacuum oven and then hybridized with 5'-end labeled $(CA)_{15}$ oligo probe. Dried gels or membranes were autoradiographed at -70° C with or without an intensifying screen.

Multiplex PCR was carried out during genotyping D11S527 & D11S534 loci through the entire 40 CEPH families. The same PCR conditions as mentioned above were applied, except for the following: 1) the reaction contained 5 pmole of primers 92005 and 91249 (D11S534) and 2.5 pmole of primers jg4A and jg4C (D11S527), 2) the mixtures were aliquoted into 96-well format PCR plates (Perkin Elmer/Cetus 9600R), 3) thirty cycles of amplification were carried out by using the "touchdown" program with annealing temperatures decreasing by 1° C in every cycle for the first 10 cycles (Don et al. 1991). Amplified products were resolved on formamide/urea gel and transferred onto Hybond N+ membrane. For detecting one of the two loci, the membranes were first hybridized with the appropriate locus specific oligo (LSO). After autoradiography, the membranes were stripped and

rehybridized with the LSO corresponding to the second locus.

9. Chromosomal Sublocalization

The subregional localization of the microsatellite markers was achieved by one or more of following methods.

a. Chromosome 11 specific somatic cell hybrid panel.

We have four chromosome 11 specific somatic cell hybrids containing different regions of chromosome 11 on the mouse background (Maslen et al. 1988)(Fig.I-3). PCR of the DNAs from those cell lines by using locus specific PCR primers would directly map the locus on the intervals of chromosome 11q.

b. Linkage analysis in CEPH families.

We obtained DNAs from 40 reference families through CEPH (Centre d'Etude du Polymorphisme Humain). Those families have,in most cases, four grandparents, two parents and an average of eight children. Such families are ideal for genetic mapping because the cis-trans relationship (i.e linkage phase) of alleles in the parents can frequently be inferred from grandparental DNAs, and once this is determined, crossovers can be counted in the meioses giving rise to the children. On chromosome 11, there are 162 RFLPs markers that have already been put through the

CEPH families. This allowed pairwise linkage analysis of new markers and the markers on the preexisting genetic linkage map of chromosome 11q. Further analysis using the multipoint method will allow us to determine the order of the new marker and its closely linked markers.

c. <u>In situ</u> Hybridization.

The fluorescence <u>in situ</u> hybridization (FISH) analyses of chromosome 11q specific cosmids were performed either by Evans' group or by Nakamura's group.

10. Cosmid DNA Sequencing

Cosmid DNA sequencing was performed either by using the Sequenase kit (USB) or by using the double stranded DNA (dsDNA) cycle sequencing kit (BRL). While in the former reaction, T7 DNA polymerase was used under room temperature; in the later reaction, thermostable Taq DNA polymerase was applied at a much higher temperature. T7 DNA polymerase is ideal for DNA sequencing because it incorporates nucleotide triphosphates more efficiently. Therefore, the range of readable sequence from the sequencing primer is about 400-500 bp in plasmid DNA preps. However, a major disadvantage of this enzyme is that it is thermounstable. As the result, alkaline denaturation of dsDNA template has to be performed prior to enzymatic reaction, allowing the primer to anneal to

the single stranded DNA (ssDNA) template. Such pretreatment has given inconsistent results in cosmid sequencing, suggesting insufficient denaturation of the large sized double stranded cosmid DNA templates. Compared to T7 DNA polymerase, Taq DNA polymerase is thermostable and retains its high activity in the denaturation step which involves a repetitive series of temperature changes from 94° C to 50° C, allowing complete separation of dsDNA and counteracting the renaturation that may occur during synthesis reaction.

Cosmid cCI11-415 was isolated from a human chromosome 11q specific library (Tokino et al. 1991) and is discussed below (38). After being partially filled-in, Sau3A-end fragments were subcloned into pTZ18u SalI cloning site. A positive subclone, XH415, was isolated by colony hybridization with 5'-end labeled (CA)15 oligo probe and was then sequenced by using the Sequenase kit (USB). A (CA)24 perfect repeat was found in XH415 and PCR primers flanking this repeat were designed (primers XH415-1 and XH415-2 While PCR amplification of human genomic DNA from unrelated individuals, as well as DNA from cosmid 415, resulted in complete failure, the amplification of DNA from plasmid XH415 successfully generated the product with the correct size (data is not shown). The DNA rearrangement during the subcloning process was thought to be the cause of PCR failures. Since the cosmid containing a long run of

perfect repeat could be highly polymorphic, direct cosmid sequencing was performed to obtain enough sequence data for designing new PCR primer by using a cycle sequencing kit (BRL) with PCR primers XH415-1 and XH415-2.

Cosmid 9/4 and 1/16 were isolated from a different human chromosome 11q library (Lichter et al. 1990) and are discussed below (ρ38). The same conditions used for subcloning, screening and sequencing cosmid cCI11-415 were applied to cosmid 9/4. Further characterization of one positive subclone, 9/4-1, revealed that it contains a (GT)21 repeat. Therefore PCR primers proximal to the target sequence were designed. Both the PCR amplification of DNA from cosmid 9/4 and from plasmid 9/4-1 were successful, exhibiting the correct size bands on the ethidium bromide stained agarose gel. However, PCR of human genomic DNA from unrelated individuals showed no amplified product. There is no indication of errors, such as a vector sequence being included in the primer sequence, in these PCR primer sequences. Likewise, cosmid 1/16 was digested with AluI and subcloned into the AluI cloning site of vector pTZ18u. One of the subclones, 1/16-2, showed a positive signal in colony hybridization and was then isolated and sequenced. As shown in the appendix, one of its cloning sites (AGCT) is too close to the target sequence (GT)14 to design a PCR primer. Thus, the primer (#22314) at the 3′ flanking region was designed, synthesized and then used as the

sequencing primer in the cosmid sequencing, using the Sequenase kit (USB).

11. Detection of New Mutation of Microsatellite Markers

Three new mutations were found in genotyping loci D11S534 and D11S439 through the entire 40 CEPH families. There is a total of 530 individuals in the 40 CEPH families, including 337 offspring who account for 674 possible informative meioses. In addition, the true parentage of these individuals has been well tested and confirmed by extensive genotyping analysis with RFLPs, VNTRs and microsatellite DNA markers in numerous laboratories around the world. The assessment of the size of each allele was based on a sequencing ladder of known sequence. To exclude possible sample mix-up and PCR artifact, the DNA samples for these individuals and their parents were obtained from three different aliquots , including one freshly made aliquot and were amplified by at least three independent polymerase chain reactions. For one of the loci, D11S534, multiplex PCR was performed to simultaneously amplify D11S534 and D11S527 loci. The genotyping data at the D11S527 locus served as the internal control for determining the new mutation at the D11S534 locus. Further, the PCR products were resolved on either 6% denaturing gel or 32% formamide plus 8M urea gel to exclude possible incomplete DNA denaturation. All of these trials gave consistent results.

Results

Screening

Since I began to characterize 11q microsatellite markers, the screening protocol has been improved achieve higher specificity and simpler procedures. The length of time for hybridization of nitrocellulose filter or Hybond N^+ membrane with $(CA)_{15}$ oligo probe has been decreased from 3hr to 1hr and the temperature of hybridization increased by 10° C, at 60° C. In addition, the washing condition is more stringent, using 0.1xSSC/0.1%SDS at 65° C ,6-8 min, for the final wash instead of 6xSSC/0.1%SDS. In some experiments, internal control utilizing (CA)n (n>12) containing cosmids were set up with cosmids that needed to be screened. The judgement of signal intensity was based on an arbitrary scale in the earlier experiments or by comparison with control cosmids in the later experiments. The total number of 50 chromosome 11q specific cosmids from our two collaborators Drs. G. Evans and Y. Nakamura, were screened with 5'-end labeled $(CA)_{15}$ oligo probe (Fig. I-4). Those cosmids were selected for screening by virtue of being assigned to the gap regions on the chromosome 11q index map, including 11q14, 11q22-q23 and 11q24-q25. This was accomplished by using somatic cell hybrid panels, FISH and the linkage analysis in CEPH families (Lichter et al. 1990; Tokino et al.

1991; Hori et al. 1992). The majority of cosmids screened were from the 11q24-q25 region (Table I-1), and they represented 38% (19/50) of the total 11q cosmids. Seven (14%) were from the 11q22-q23 region, seven (14%) from 11q13, two (4%) from 11q14 and two (4%) from 11q21. The remaining cosmids were assigned grossly to the long arm of chromosome 11. Of those 50 11q cosmids, fourteen (28%) gave strong signals and were then subjected to further investigations; seven (14%) gave weaker signals and 4 of these 7 cosmids were further studied to insure that new markers in gap regions were not missed. Three cosmids showing intense signals (cosmids 2G9, 8G11 and 9A2) were known to be overlapped and were scored as one cosmid (Evans, personal communication). Thus, in fact, the total number of cosmids giving stronger signals were 12. The results of Southern blotting on 16 (CA)n positive cosmids (strong and weaker signals) revealed that 13 (81%) of them had a single copy of poly CA repeat in their inserts; the others (19%) had more than one repeat block in their inserts.

Subcloning and Sequencing of (CA)n Positive Clones

All 12 cosmids showing strong signals were subcloned and sequenced. Eight of those cosmids (67%) contained a single perfect repeat, three (25%) had two separate perfect and/or compond perfect repeats, one (8%) had an interrupted repeat. Cosmid 1/16 had a (GT)14 repeat too

close to the cloning site to design a PCR primer and was then sequenced directly off the cosmid template to get enough sequence data (see "Direct Cosmid Sequencing" for more information).

Polymerase Chain Reaction and Genotyping of Microsatellite Markers through 40 CEPH Families

Sets of oligodeoxyribonucleotide primers for PCR were designed for cosmids with consideration of the sequences being unique, close to the target (CA)n block and within a narrow range of Tm. Although various factors were considered in designing PCR primers, we still had some PCR failures. For example, PCR amplification of locus DRD2 resulted in two major set of bands or "doublets" approximately 20 bp apart. Efforts to eliminate the "doublets" failed, including using higher electrophoresis temperature. By examining more carefully the flanking DNA sequence around DRD2 locus, we have found that a 20 bp long segment is located 20 bp from the target sequence of DRD2 which shares 85% (17/20) of homology with the 3' PCR primer. A new 3' primer was then designed and the doublet completely eliminated.

PCR amplifications of DNA samples from a number of unrelated individuals were carried out for detection of polymorphisms. Following the PCR, the amplified products were analyzed on ethidium bromide stained 1.5-2% agarose gels to insure a sufficient amount of the correct size

products had been obtained. Once the correct PCR products were detected, a larger scale of PCR on DNA samples from about 20 CEPH parents was performed.

As pointed out in Materials and Methods, there are three ways of visualizing the PCR products on the denaturing gel: 1) internally labeling the products with α-32p-dCTP during DNA synthesis by Taq polymerase, 2) 5'-end labeling of one of the primers, 3) capillary blotting of products onto the membrane and then hybridize the membrane with labeled $(CA)_{15}$ oligo probe.

In the early stage of characterizing the 11q microsatellite markers, the internal labeling method was the first choice because it is simpler than either end labeling or blotting techniques. An example of genotyping a CEPH family using this method is shown in figure I-5A. It shows that while a pair of the same alleles from a homozygote gives 2 discrete, equally intense bands, two alleles from heterozygotes give 4 bands. It has been demonstrated that of those 2 bands from a homozygote, the faster moving band is actually the CA strand, whereas the slower moving band is the GT strand (Weber and May 1989). Sometimes this effect can pose the problem in genotyping randoms when 2 alleles are very similar in size. The second most commonly used technique for detecting the products on the gel is to label one of the PCR primers. In theory, only one of the strands (either GT or CA strand) will be labeled by

polymerase chain reaction. However, after resolving on the gel, each allele gives one major band as well as one to four extra bands which are less intense. These extra bamds are two nucleotides or multiples of two nucleotides shorter than the major band. Such examples are shown in figure I-6. Extra bands have been a problem for genotyping microsatellite markers not only in our lab but in many others which are involved in developing or utilizing (CA/GT) microsatellite markers. One way to reduce the background and eliminate the extra bands is to resolve the PCR products on a 6% polyacrylamide, 32% formamide plus 5.6 M urea gel, blot the products onto Hybond N^+ membrane, and then hybridize the membrane with labeled $(CA)_{10}$ or $(CA)_{15}$ probe. It has been demonstrated that, in many instances, this method is simple and powerful.

A total number of 14 11q subclones containing a longer run of CA repeat units (12) and shorter repeats (2) have been tested in random for detecting the presence of polymorphism. Nine of them are polymorphic with heterozygosity or PIC varying from .31 to .88 and the number of alleles range from 2 to 11. Five of them are index quality markers with heterozygosity or PIC value greater than .70. Those results are summarized in Table I-2 and I-3. It was surprising that at least 4 subclones which contain a long run of repeat units are not polymorphic at all. When we tested the correlation

between the length of repeat units and the degree of polymorphism in all subclones containing (CA)n repeats, we were unable to find a positive correlation between these two parameters (Fig. I-7). The correlation coefficient (r) calculated from this set of data was -0.27. We then tested the correlation between the number of repeat unit and the degree of polymorphism in 9 polymorphic markers (Fig. I-8). The correlation coefficient is 0.89. While the former test failed to show the correlation of repeat length vs polymorphism in all (CA)n containing subclones, the latter test indicates that the 2 parameters have positive correlations.

We tested one non-polymorphic poly CA repeat cosmid 2/11, which contained a perfect (CA)23 repeat block, with rodent genomic DNAs. The DNA samples from human, mouse and hamster were amplified under the same condition as mentioned in the "Materials and Methods", except that annealing temperature was reduced to 53° C, for 50 sec. A control sample containing all reagents except DNA was also amplified at the same time to insure that no contamination occurred during the preparation. SIgnals from PCR products with appropriate intensity and size were seen in both mouse and hamster lanes on the X-ray film. When compared with humans, the hamster had the same allele, whereas the mouse had one one allele of the same size and one different allele (Fig. I-5B). (No products could be detected from the control tube.)

All five index quality markers (DRD2, D11S490, D11S527, D11534 and D11439) were genotyped through the entire 40 CEPH families (typing of DRD2 and D11S490 was done by Jeff Luty, a former technician in the lab). Two of them, D11S527 and D11S534, were typed simultaneously by using a multiplex PCR technique. At the equimolar primer concentration for both D11S527 and D11S534 loci, the locus D11S527 was amplified much more efficiently and gave much stronger signals on the autoradiograph film compared with those of D11S534, probably because it has an approximately 80 bp shorter target sequence. The primer concentration for D11S527 in the reaction was, therefore, decreased by 75%. In addition, a "touch-down" PCR program was performed to compromise annealing temperatures which are optimal for each locus. This allowed successful amplifications of both loci. In addition to using multiplex PCR, multiple gel loading technique was also applied in the typing process. This allowed us to load several different samples in one well in the sequencing gel at 20 minute intervals. Figure I-9A shows that PCR products loaded twice with 20 min interval between loadings and resolved on a denaturing gel without formamide. The products were transferred onto Hybond N^+ membrane by blotting and then hybridized with a labeled locus specific oligo probe. After typing one locus, the membranes were stripped and hybridized with the second locus specific oligo probe (Fig.I-9C).

Consequently, on a single denaturing gel, as many as four loadings can be applied. This allows scoring of a large number of CEPH families with one gel run. This method has proven to be very efficient, reliable and powerful. The background seen in figure I-9A, an occasional problem in the scoring, has been substantially reduced by using formamide-urea gel as shown in figure I-9B.

Direct Sequencing of Cosmids 1/16, cCI11-415 and 9/4

Direct cosmid sequencing, either by using the Sequenase kit (USB) (which requires denatured DNA) or a dsDNA cycle sequencing kit (BRL), was successful with cosmids 1/16, cCI11-415 and 9/4 following slight modifications. One example of cosmid sequencing is shown in Fig.I-10. In the case of cosmid 1/16, it provided the unambiguous sequence information for designing the primer (#41919), which is located 5' beyond the AluI cloning site present in the plasmid 1/16-2. Thus, re-subcloning, re-screening and sequencing procedures, which can be time consuming and tedious processes, became unnecessary. PCR amplification by using primers # 22314 and #41919 on a number of unrelated individuals, revealed 9 allelic polymorphisms with heterozygosity of 0.72.

To determine the artifactual recombination site, target DNA sequence from unrearranged cosmid cCI11-415 was compared with that of the rearranged plasmid XH415. As shown in Fig.I-11, the recombination site is

undoubtedly at the nucleotide 98, since both cosmid cCI11-415 and plasmid XH415 share the identical sequence from nucleotide 98 to nucleotide 437, but have completely different sequence from the 1st nucleotide to the 97th nucleotide. As suspected, the recombination site in the plasmid XH415 was located within the target sequence so that one of the PCR primers was designed in the 5′ upstream region from the breakpoint and another primer from the 3′ downstream region, causing the PCR failure. There are two interesting features in this 437 bp region. First, the rearrangement occurred within a short alternating purine-pyrimidine tract, which is partially homologous between cosmid cCI11-415 and plasmid XH415, presumably functioning as a starting point for homologous sequence paring. Second, the DNA sequence flanking the breakpoint was very rich in short simple-sequence tandem repeat tracts, such as (CA)3, (CA)4, (GA)3, (GA)4 and (CAGA)2. Such sequences can exist as left-handed or Z-DNA conformation, and have been proposed to represent regions of altered chromatin, and thus altered chromosome structure. Meanwhile, direct cosmid sequencing by using primer XH415-1 revealed the similar sequence motif at its 3′ extended region where short tandem repeats are heavily clustered (Fig.I-12). The third PCR primer, XH415-3, was designed according to the correct cosmid sequence at the 5′ flanking region of the (CA)24 repeat. The PCR

amplification of DNAs from a number of randoms revealed monomorphism.

Direct cosmid sequencing on 9/4 showed the identical sequence between cosmid 9/4 and plasmid 9/4-1 in a few hundred base pair long region, which surrounds the $(GT)_{21}$ repeat. Although we were unable to identify the rearranged sequence, we believe that the rearrangement breakpoint is within $(GT)_{21}$ repeat region, because the sequences of both PCR primers are unique and close to the target sequence (Fig.I-13). Indeed, a high resolution FISH study carried out by Lichter et al. suggests the idea that chromosomal rearrangement may have occurred to cosmid 9/4 during the genomic DNA cloning process, since it hybridized to two discrete sites at band q23 on the long arm of human chromosome 11 (Lichter et al. 1990).

New Mutations Found at D11S534 and D11S439

Individual 137701, the son of 137710 (father) and 137711 (mother), is a heterozygote at the D11S534 locus. His father's genotype is A4,A6 and his mother is A2,A4. While he inherited allele A6 from his father, his other allele is inconsistent with either alleles of his mother. It is two nucleotides shorter than allele A2 but two nucleotides longer than allele A4 (Fig.I-15). Furthermore, the new mutation found in individual 137701 is inheritable since it transmitted to four of his offspring in Mendelian fashion. Likewise, three

individuals from CEPH family K12 were re-examined in three independent PCRs. The result has been consistent. Both parents, 1201 and 1202, share the same alleles A3 and A4. Individual 1204 inherited allele A3 from one of her parents. Her other allele apparently is a new mutation which is two nucleotides longer than allele A3 (Fig.I-15). There was no mosaicism detected in these two new mutations, indicating that they most likelyare germline mutations. The third mutation was found in family K102 at the D11S439 locus. Individual 10207, the daughter of 10201 (father) and 10202(mother), has three alleles A2,A9 and A8 (data not shown). She inherited the A2 allele from her father who has A2/A10 alleles and A9 from her mother who has A2/A9 alleles. Her extra allele A8, which is two nucleotides longer than allele A9, is apparently a new mutation. An repeat experiment using a new aliquot of DNA sample gave the same result.

Discussion

Characterization of highly polymorphic microsatellite markers from chromosome 11q.

The estimation of the number of (CA)n repeat with "n" greater than 12 in the human genome ranges from about 35,000 to 130,000 by using hybridization method (Weber, 1990). This may be an over-estimated figure, since colony screening or genomic DNA hybridization with (CA)n oligo probes have the potential of picking up short repeat units as well as medium length, but interrupted, repeat units; those shorter repeats and interrupted repeats are often excluded from the final results by the time data are compiled for publications. Further, without sequencing data, hybridization results should be considered as very preliminary data.

The results of screening chromosome 11q cosmids, as reported here, have several interesting implications. First, in this study, of 50 11q cosmids screened and consequently sequenced, fourteen (28%) showed strong signals when hybridized with 5'-end labeled $(CA)_{15}$ oligo probe, suggesting that one (CA)n (n>12) repeat appears in about 140 Kb human genomic DNA sequences. (The insert size of those 11q cosmids ranges from 35 to 40 Kb.) Thus, this indicates that there are probably a total of 2.4×10^4 copies of (CA)n repeats with "n" greater than 12

in the haploid human genome, 1063 copies on the human chromosome 11 (144 Mb) and 627 copies on the long arm of chromosome 11 (86 Mb) (since chromosome 11 represents 4.51% of the genome, Morton, 1991).

Second, in this analysis, we screened about an equal number of 11q specific cosmids which were isolated from the 11q24-q25 region (19/37) and from regions other than 11q24-q25 (18/37). While about one half of the 18 cosmids from other regions contained poly CA repeats, only one fifth of the 19 cosmids from 11q24-q25 contained poly CA repeats, suggesting the non-random distribution of (CA)n repeat along the chromosome 11q. Non-random distribution of (CA)n repeat may pose a problem in the future isolation of polymorphic microsatellite markers from the very distal region of chromosome 11q, where very few index quality markers have been isolated. Therefore, Chi Square(x^2) statistical analysis was performed to test the significance of the difference in the occurrence of (CA)n repeat between two groups (11q24-q25 region and other regions) (Table I-4). There was no significant difference found between those two groups (α=0.25), indicating that the distribution of (CA)n repeats is probably random along the 11q. It has also been demonstrated that the poly CA repeats are equally distributed in both CpG-rich and non-CpG-rich regions along the long arm of chromosome 11 (Hermanson et al. 1992). These data are consistent with the previous

observations that simple-sequence tandem repeats are randomly distributed in the human genome.

Third, the data compiled from 12 (CA)n repeats with "n" greater than 12 failed to demonstrate the correlation between the number of repeat units and the degree of the polymorphism (r=-0.27), which has been observed by Weber (Weber,1990). However, we did find the positive correlation (r=0.89) between the percent of heterozygosity and the number of repeat units as long as the locus showed polymorphism using our very limited samples. others have also reported similar results by using a much larger data set. Hudson et al. (19920 reported 136 human (CA)n repeats with n varied from 13 to 28. Their data showed much less linear correlation between the percent of heterozygosity and the repeat length if the repeat units were less than 23. However, the percent of heterozygosity increses with the repeat length if the repeat units were greater than 23. These data suggest that the chromosomal location of a repeat may play a role in the presence or absence of polymorphism.

In addition, we also present an example of genotyping multiple loci simultaneously through 40 CEPH families by utilizing multiplex PCR, multiple loading, blotting and sequential hybridization techniques. The beauty of combined methods as presented here can be summarized as follows: 1) It has potential to maximally

save time, labor and PCR reagents, especially expensive Taq polymerase. 2) The PCR products from at least 10 CEPH families can be resolved on a single denaturing gel, blotted on one membrane and stripped repeatedly many times (I have reused membranes for more then 6 times and they still gave very good signals.) Thus, genotypes for each locus on all CEPH family members can be stored on one membrane for a long period of time and retrieved at any time if necessary. 3) The "cold" PCR products can be stored in 4° C for several months without the degradation, unlike radioactively labeled products. 4) The confirmation or exclusion of new mutations, sample mix-up, mistyping and non-paternity can be made more efficiently and precisely, since the two loci amplified simultaneously will serve mutually as internal controls. 5) No expensive equipment, such as an electroblotting apparatus, is needed.

We have observed the conservation between human and rodent at one of the non-polymorphic (CA)n repeats isolated by us. The short stretch of unique sequences flanking the repeat is probably identical between man and rodent. While the hamster shares the same allele as humans, the mouse exhibits two allelic polymorphisms. Similar conservation of the unique sequences flanking the simple sequence tandem repeats has also been found between man and primates at a highly polymorphic microsatellite marker locus D11S527, where DNA samples

from a number of primates showed much shorter repeat tracts compared with that of humans (David Browne, personal communication). Since the majority of simple sequence tandem repeats are probably non-functional, therefore, not under selective pressure, they are more likely to diverge over 70 million years after the divergence of rodent and man. This is exactly what we have found in cosmid 2/11 in which the unique, flanking sequences are well conserved, but the dinucleotide repeat region has undergone changes. Such changes are most likely to be the deletion or insertion of one or more repeat units from the original tract by slipped-strand mispairing mechanism (Levinson and Gutman 1987b).

Recombinational breakpoints in cosmids cCI11-415 and 9/4.

We have successfully sequenced three 11q cosmids which allows us to locate precisely one DNA rearrangement breakpoint found in cosmid cCI11-415 and speed up the process of isolation and characterization of (CA)n microsatellite marker locus D11S490 contained in cosmid 1/16. Our results presented here may have 2 interesting implications. First, two cases of DNA rearrangements share similarity in that they both contain alternating purine-pyrimidine repeats in close vicinity of the breakpoint (cosmid cCI11-415) or putative breakpoint (cosmid 9/4), although the rearrangements occurred

clearly at different levels; it occurred at the genomic DNA level for cosmid 9/4, but at the cosmid DNA level for cosmid cCI11-415. There are two possible ways to generate the rearranged DNA, as we have found in cosmid cCI11-415. 1) It could be caused by reannealing of DNA fragments by T4 DNA ligase during ligation reaction, which would bring two noncontiguous fragments into juxtaposition. 2) A rearrangement could be caused by unequal crossing-over, deletion and insertion. The rearrangement caused by reannealing can be detected unambiguously by identifying the restriction enzyme recognition sequence at the junction of those two fragments. There is, however, no evidence of presence of a Sau3A recognition sequence at the breakpoint, indicating strongly that the rearrangement could be caused by a special feature of the DNA sequence flanking the breakpoint. Indeed, we have found numerous short alternating purine-pyrimidine tracts very close to the breakpoint. More significantly, we have identified a CACATC tract at the breakpoint which is partially shared by unrearranged cosmid and rearranged plasmid DNAs, presumably initiating the DNA rearrangement. The hypothetical event which occurred in cosmid cCI11-415 is illustrated in Fig.I-14. Although we were unable to identify the exact location of the breakpoint in cosmid 9/4, results from high resolution FISH done by Lichter et al. and from our own polymerase chain reactions suggest

that the breakpoint could perhaps be right in the (CA/GT)21 repeat region of cosmid 9/4.

Second, it is known that human chromosome 11q contains a number of sites of chromosome rearrangement involved in tumorigenesis, including t(11;22)(q13;q13) in B-cell chronic lymphocytic leukemia, B-cell Non-Hodgkin lymphoma and multiple myeloma; t(4;11)(q21;q23) in infantile acute lymphoblastic leukemia; t(9;11)(p22;q23), t(11;19)(q23;p13) in acute monocytic leukemia; and t(11;22)(q24;q12) in Ewing sarcoma.

Both cosmids 9/4 and cCI11-415 were isolated from this distal region. More significantly, the cosmid 9/4 has been localized right in the flanking region of the chromosome breakpoint found in t(4;11)(q21;q23) and t(11;19)(q23;p13) (Selleri et al. 1991). It would be interesting to know whether the recombinogenic properties exhibited by these cosmids reflect the DNA instability at certain regions on the distal 11q.

Cosmids 9/4 and cCI11-415 have been localized to band 11q23 and 11q24-25, respectively. More significantly, cosmid 9/4 is right in the flanking region of the chromosome breakpoint found in t(4;11)(q21;q23) and t(11;19)(q23;p13) (Selleri et al. 1991).

Spontaneous new mutation of microsatellite markers

It has been estimated that new mutations of poly CA and other simple-sequence repeats are low, about $1X10^{-4}$

per chromosome per generation (Weber 1990). In this report, we present further data on new mutation rates of 3 microsatellite markers, D11S534 D11S527, and D11S439, on human chromosome 11q. The data from the D11S534 locus indicate that the mutation rate of poly CA repeat ranges from 0 to 6.1×10^{-3} per chromosome per generation, which is about 30 times higher than the previous estimation.

The type of mutation that we have observed includes both germline and somatic mutations. Since the new mutation found in individual 137701 has been transmitted to four of his offspring in Mendelian fashion and no mosaicism has been detected, it is most likely to be a germline mutation. Likewise, the mutation found in 1204 is most likely a germline mutation because no mosaicism has been detected. On the other hand, the mutation found in 10207 is probably a somatic mutation. According to the intensity of alleles, it is possible that her cells consist of two cell lines observed at 50 percent each.

We see no bias towards longer repeats, since we have not yet detected a single new mutation at the D11S527 locus in the 40 CEPH families. This locus contains not only a longer run but also a compound repeat. In VNTR markers, the mutation rate increases with heterozygosity, as pointed out by Jefferys et al (1988). From our admittedly limited data, there is no indication that the degree of the heterozygosity affects the mutation rate in poly CA microsatellite markers.

One common feature that we have seen in all of our mutations is that mutated alleles are always two nucleotides shorter or longer than the parental alleles. While the mutations found in individuals 1204 and 10207 can be explained as insertions of a single repeat unit, the mutation found in individual 137701 can be interpreted as insertion or deletion of one unit, corresponding very well with one of features of a slipped-strand mispairing event. The mutation presented here is strongly biased toward one repeat unit which has been reported in other organisms as well, presumably requiring the least energy (Levinson and Gutman 1987a, 1987b). The deletion mutation is not favored vs insertion mutation, since two of three cases were clearly caused by insertion and one other case can not be determined because of uninformative data.

Slipped-strand mispairing (SSM) has been considered as the major mechanism for new mutations in poly CA and other simple-sequence repeats in prokaryote as well as for the new mutation of VNTR markers during spermatogenesis (Levinson and Gutman 1987a, 1987b; Jefferys et al. 1988). Slipped-strand mispairing may occur during DNA replication and repairing processes. During replication, an extending DNA strand may slip backward by one or more repeat units resulting in a single or multiple repeat unit insertion (Fig.III-1). On the other hand, if slippage occurs in the template

strand, it will end up with a deletion. Thus, the length of alleles of poly CA microsatellite markers varies by one repeat unit or multiples of one repeat unit. An exception to this was the report of where two alleles were different by 1 base pair (Wang and Weber 1992). Furthermore, the mutation rate of poly CA repeats should be high enough to generate extensive polymorphism, and low enough to transmit faithfully from generation to generation. In humans, the new mutation rate ranges from $3X10^{-3}$ to $1X10^{-4}$ per chromosome per generation which, is lower than prokaryotes (Levinson and Gutman 1987a, 1987b; Freund et al.1989); the rate does not correspond to the length of repeat region as seen in prokaryotes. Despite these slight differences, the major features of new mutation found in both prokaryote and eukaryote show strongly similarities. Therefore, we believe that the mutations described here were generated by the slipped-strand mispairing mechanism.

Section II

Construction of the index map of the long arm of human chromosome 11

Materials and Methods

DNA samples

The DNA samples of 530 individuals from 40 reference families were obtained from CEPH. Those families used were 13291, 13292, 13293, 13294, 1331, 1332, 1333, 1334, 1340, 1341, 1344, 1345, 1346, 1347, 1349, 1350, 1362, 1375, 1377, 1408, 1413, 1416, 1418, 1420, 1421, 1423, 1424, 66, 12, 23, 21, 2, 17, 37, 35, 28, 45, 884, 102, 104.

Genetic markers

Seventeen markers of index marker quality were included in the linkage analysis. Of them, seven microsatellite markers were developed by our laboratory. The procedures for isolation, characterization and genotype determination of (CA/GT)n microsatellite markers on the long arm of human chromosome 11 have been described under the materials and methods section in the Section I. For each index marker used for construction of the index map, the locus symbol, the probe name, the heterozygosity or PIC, the type of polymorphism and the related literature are listed in Table I-5.

In addition to microsatellite markers developed by us, the data of four haplotyped RFLPs, which preexisted in the CEPH database V5 as multiple probe/enzyme polymorphic systems, were provided by Mike Litt; data for markers D11S614 and Mfd231 were contributed by Jim Weber; data for markers INT2, PYGM, sMSH3, sMSH2 and CD2 were contributed by Helen Donis-Keller.

Error detecting during the linkage analysis

Genotyping data for each individual in the CEPH families were checked by at least two persons at different steps, including film reading and data entering. Inconsistent genotypes were reexamined by repeated PCR amplification and genotype determination. Further, CHROMPIC option of theCRIMAP computer program was used to flag isolated double recombinant within the short genetic distance, which has been considered to be the probable result of data errors (P. Green in preparation 1991). When CHROMPIC detected two recombinations within a 10 cM or less distance, the film and /or the genotype interpretation were reexamined. A change in the CEPH database was made only when the new data was unequivocal. In some instances, the genotype data of individuals in question at the locus were subtracted from the CEPH database, i.e., they were coded as "unknown".

Linkage analysis

The measurement of support for the linkage of a pair of loci is defined as the decimal logarithm of the ratio of the probability that the data would have arisen if the loci are linked to the probability that the data would have arisen if the loci are unlinked (the lod score). The conventional threshold for declaring linkage is a lod score greater than 3, which would indicate that the observed data are 1000-fold more likely to occur if two loci are linked than they are unlinked.

Genotype data, except the data contributed by Helen Donis-Keller, were entered into the CEPH database and prepared in the format which can be used by the CRIMAP program by using PEDOUT, a CEPH subprogram. Two sets of data, one from CEPH database and one from Helen Donis-Keller, were then incorporated into one set of data according to the family ID number and individual ID number, by using the MERGE option of CRIMAP. Two point and multipoint linkage analysis were performed by using CRIMAP V2.4 on a SUN work station computer. The TWOPOINT option of CRIMAP was used to calculate the maximum lod score (^Z) vs theta (ô) between any two loci tested. The order of index markers was generated by using the BUILD option of CRIMAP based on these two-locus linkage data. The BUILD used two lci with the greatest number of informative meiosis in the CEPH database as a starting pair. Then other loci were inserted one at a time into

the order. The following three measures were taken to insure that the correct order was accepted. 1) The likelihood of 1000:1 odds was required for each locus to be incorporated into the map. 2) All index markers were subdivided into several subsets and each subset was used as a starting core where other markers were added on to it. The current index map was ultimately merged from those overlapping maps. 3) The FLIPsn option of CRIMAP was used to generate the permutations of three adjacent loci at a time, after the order of loci had been generated by BUILD. In addition, the maximum LOD score (^Z) vs theta (ô) between any two loci were arranged in an easy-read format—"road map" format by using the TWOTABLE option of the CRIMAP program. The genetic distance was calculated by using the Kosambi mapping function.

Results

Index markers

Several features of 17 markers used in this linkage analysis are summarized in Table I-5. These markers show three striking features. First, they are highly informative with heterozygosity or PIC > 0.70. Second, most of these index markers can be very efficiently genotyped through all CEPH families by multiplex PCR amplification and blotting methods. Third, these PCRable

markers are STS (Sequence Tagged Site) markers which can be used as anchor points in physical mapping of human chromosome 11.

Linkage analysis and error filtration

Procedures of two-point and multipoint linkage analysis, by using TWOPOINT, TWOTABLE, BUILD and FLIPsn options of the CRIMAP program, have been described under materials and methods. Two point linkage data is given in Table I-6. We were able to uniquely place 14 index markers on the map with odds exceeding 10,000:1. However, we were unable to uniquely place marker loci D11S527 and CD2 on the map because they have zero and 2 per cent recombinations with marker loci D11S533 and D11S614, respectively. The index map of the long arm of human chromosome 11 was anchored by marker loci D11S288/D11S436. They have been localized on the short arm of chromosome 11 and are 4.1 cM away from sMSH3, the most proximal locus on the long arm. We further detected the potential data errors in the database, which were shown as isolated double recombinants, by using CHROMPIC. Six unequivocal isolated double recombinants were found in the CHROMPIC output file as shown in Table I-7. After eliminating those possible errors from the databases, we performed the multipoint linkage analysis again by utilizing BUILD and FLIPsn options. We have found that there is no difference between two orders of 15 index markers before and after error filtration. There is,

however, 3% decrease in the length of the map after error filtration. The order of 15 index markers is shown in Fig.I-16, with a confidence level of 10,000:1 in support of it being the best order. The sex-averaged index map of the long arm of human chromosome 11 spans 82 cM from D11S288/D11S436 to D11S439. The sex-specific index maps span 118 cM and 53.4 cM for female and male, respectively (Fig.I-17). The genetic distance between two adjacent index markers is varied from 1.4 cM to 14.7 cM, which meets the requirement for the index markers.

Table I-7. Isolated double recombinants

Individual	Markers
133204	301/PCR
133408	6d11/pcr
134004	4F7/PCR
134707	S144/S351/HAPLO
1329309	sHI-1
140814	INT2

Discussion

The first linkage map of human chromosome 11 was reported by Leppert et al (1987). It was followed by a more detailed map of the long arm of chromosome 11 constructed by Julier et al. (1990). The latter map consists of 31 RFLPs or VNTR markers, but less than 7% of those markers have herterozygosities greater than .70. Since most markers used were not very informative, the odds against the inversion of orders of 10 loci on the

map were less than 100:1, making it much less reliable in certain regions for physical mapping, positional cloning, etc. Additionally, these marker loci were clustered in two chromosome bands, 11q13 and 11q23.

Compared with preexisting chromosome 11q maps, the map, as we presented here, has shown two significant improvements. First, seventeen markers used are highly polymorphic with heterozygosity greater than .70. Most of them are simple sequence repeat microsatellite markers and can be genotyped efficiently through all CEPH families by using multiple PCR and blotting procedures, as we described in this thesis and as reported by others. It has been demonstrated that 80% of microsatellite markers will have heterozygosity or PIC >.50 and about 50% will have heterozygosity or PIC >.70. Therefore, it is clear that to fulfill the goal of constructing an index map and ultimately the construction of a high-resolution map of the human genome, developing new microsatellite markers, either from the gap region or from preexisting RFLP markers, is the best option. It is perhaps worth mentioning that a recent report by Iizuka et al. (1992) described isolation and characterization of 7 Alu repeat polymorphic markers from human chromosome 11. Unfortunately, there are two major drawbacks of Alu repeat polymorphic markers: 1) low heterozygosity (none of the 7 markers have heterozygosity or PIC exceeding .60), 2) much longer genotyping process. Since the SSCP

procedure requires very low voltage during electrophoresis, it considerably prolongs the time for allele resolving.

The second improvement of the index map is that fifteen index markers are evenly distributed along the map with interlocus distances of less than 15 cM. The sex-averaged linkage index map of the chromosome 11q extends 82 cM from D11S288/D11S436 to D11S439, which covers the entire long arm except the very distal region. The confidence level of accepting the order presented herein is 10,000:1, and odds against inversion of all marker loci by permutating 3 loci at a time are greater than 10,000:1 (except one pair of loci), indicating high informativeness of markers and strong support for the order.

As has been known for many other human linkage maps, the female map (118 cM) of chromosome 11q is considerably longer than the male map (53.4 cM). It has been demonstrated that the ratio of female/male map length in intervals between two adjacent markers varies from less than 1 to 6-7. The ratio tends to increase from centromere to telomere along the chromosome arms (Hazan et al. 1992; Wang and Weber 1992). In this study, we found that the ratio of female to male map length in the intervals close to the centromere are consistently about 2-3. The largest difference between the female and male maps occurs in the most distal interval, where the female

map is about 5 times longer than the male map. Only in the interval bracketed by D11S288/D11S436 and sMSH3, D11S144/D11S351 and D11S490 is the female map shorter than the male map. Morton pointed out that the genetic arm lengths of chromosome 11q for male and female are 63 and 100 cM, based on the chiasmata counts from spermatogenesis and multiple pairwise linkage analysis. Since the chiasma data for female are not available, genetic arm lengths for females (100 cM) are less reliable than for males (63 cM). Nonetheless, our result is in fair agreement with Morton's.

The CHROMPIC option of the CHIMAP program allows us to detect isolated double recombinants in the data used for the linkage analysis. Several such recombinants were identified as shown in Table I-7. As pointed out by K. Buetow (1991), most isolated double recombinants within short genetic distance are false recombinants which have been introduced into data by misclassification of nonrecombinant gametes as multiple recombinants. Often, such events occur when flanking markers are highly informative. It has been shown that such events will cause the inflation of interlocus distance, and in the construction of high-resolution linkage map, reduce the power to discriminate orders. It has been shown that the error rate of genotyping human polymorphic markers varies from 0.1% to 0.6% in different laboratories (Buetow 1991). In this study, we performed the error filtration,

and, as a result, the length of thesex-averaged map decreased by 3 per cent. In addition, we compared several map intervals shared by maps constructed by Julier et al. and by us, assuming that map inflation of subset loci within those intervals would change the overall distance. These intervals are bracketed by marker loci D11S288/D11S436, PYGM, INT2, D11S388 and D11S35. We were unable to find differences greater than 1 cM in each interval tested between two maps, indicating that our map distance is fairly accurate.

Since over half of the cosmids screened are known to be from the 11q24-25 region, the question as to why the distal marker is under-represented in our 11q specific cosmid collection arose during the experiments. There are at least two answers which are appealing to us. 1) The distribution of (CA)n repeats with n>12 is non-random, since their number on the terminal region is low. 2) With the increase of recombination frequency at the telomeric region, 1 unit map distance may well represent less than 1 megabase, which would inflate the linkage map with widely spaced markers (Rouyer et al. 1990; Morton 1991; Hazan, J. et al. 1992).

As mentioned earlier, the index map covers the long arm of chromosome 11q except the very distal region, because the D11S439 locus is known to be proximal to the most distal RFLP marker on the preexisting linkage map. Several estimated figures for the sex-averaged map of

chromosome 11q have been reported, including physical (86 Mb) and linkage (86 cM; 105 cM) maps (Morton 1991; Julier et al. 1990). It seems appropriate to compare our male map (53.4 cM) with that estimated by Morton (63 cM), since his estimation of the physical map was based solely on the chiasmata counts from spermatogenesis and to compare our sex-averaged map with that published by Julier et al., since both maps share the same anchor point. It is clear that our map is in fair agreement with both Morton's and Julier's. There is, therefore, about a 10 to 20 cM gap left at the telomere region of chromosome 11q, where one or two index markers are still needed to fill the gap in order to complete the index map.

The index map of the long arm of human chromosome 11 presented here is very useful for positional cloning of diseases localized grossly on the 11q, including multiple endocrine neoplasia type I, ataxia telangiectasia, tuberous sclerosis type II and other genetic diseases. It is also useful in integrating the physical, cytogenetic and linkage maps of the chromosome 11q either by assisting arrangement, orientation of cosmids or yeast artificial chromosome (YAC) contigs.

The sequences for designing PCR primers of 7 probes have been deposited in Genbank. The genotypes have been entered in CEPH database by technicians or students from our laboratory.

Part III

Further evidence that the slipped-strand mispairing may give rise to the extra bands

Introduction

Anomalous PCR products are seen after the PCR products are resolved on the denaturing gel. These anomalous products appear as ladder-like extra bands, which are two nucleotides apart and can sometimes interfere with the genotyping process. An example of such interfering is shown in Fig.I-6A. There have been at least five mechanisms postulated as the cause of these extra bands.

The first mechanism is slipped-strand mispairing (SSM) during polymerase chain reaction (PCR) (Litt and Luty 1989; Tautz 1989) (Fig. III-1). Luty et al. cloned the PCR products from the DXS425 locus into an E. coli strain pTZ18u. Five out of 6 subclones sequenced had 14,17,19,22 and 23 CA/GT repeat units. They proposed that such variants were generated by the slipped-strand mispairing (SSM) mechanism during the PCR. This explanation is not very convincing because it is known that synthesized (CA)n oligomers or (CA/TG)n repeat sequence generate a wide range of variable number of repeats or new mutations after being introduced into E.coli strains (Levinson and Gutman, 1987a, 1987b; Freund

et al.1989). However, it is worth investigating further by using different approaches.

The second mechanism is due to the template-independent addition of one nucleotide at the 3' end of newly synthesized DNA strands by Taq DNA polymerase (Weber 1989). This mechanism is unlikely to be the major cause of extra bands, since in most cases, the space between two ladders is two nucleotides apart, not one nucleotide.

The third mechanism is due to "out of register" annealing of truncated products (Odelberg et al. personal communication at the Cold Spring Harbor meeting 1992). Such recombinational events have been observed during polymerase chain reaction (Shuldiner et al. 1989; Meyerhans et al. 1990; Marton et al. 1991; Rajendra Kirshnan et al. 1991). In such an event, incompletely elongated DNA strands are generated. Those heterogenous PCR products then partially reanneal among themselves upon shared homologous sequences during denaturation and renaturation steps of the thermal cycles (Fig.III-2).

The fourth mechanism is due to template-switching during DNA synthesis (Odelberg et al. personal communication 1992). Instead of reannealing between themselves, truncated PCR products bind to completely different sites on the human genomic DNA via partial homologous sequences.

The last mechanism, which is irrelevant to the process of polymerase chain reaction, is due to incomplete denaturation of DNA fragments during electrophoresis. Consequently, DNA molecules with the same length and sequence will not comigrate on the gel.

In this study, we present another approach to address the question of the cause of extra bands. Since subcloning either synthesized (CA)n repeats or PCR amplified CA/GT repeat units into E.coli host strains can generate deletion or insertion in the original repeat fragments (Levinson and Gutman, 1987; Freund et al. 1989; Wilkie and Higgs, 1992; Iwasaki et al. 1992), we decided to directly sequence one of our microsatellite marker loci, which contains a compound perfect repeat segment from the PCR amplified genomic DNA samples of unrelated individuals. These individuals have been proved to be homozygotes at the locus tested. Sequence data from two individuals were obtained from both DNA strands by using one PCR primer at a time. In the control experiment, a subclone containing no CA/GT or other kinds of simple repeat sequences was PCR amplified and sequenced by the same methods. Our result strongly indicates that extra bands with two nucleotide spacing seen on the denaturing gel of dinucleotide repeat fragments are primarily due to slipped-strand mispairing mechanism. The slipped-strand mispairing preferentially occurs at certain target sequence, i.e. CA/GT dinucleotide repeat sequences. In

addition, we also investigated the molecular basis of polymorphism found at the D11S527 locus, where both tetra- and dinucleotide repeats are present. Our analysis suggests that 3/4 of polymorphism seen at this locus are due to the variations within (CA)n repeat and the rest are due to the variations in both tetra- and dinucleotide repeats.

Material and Methods

Dinucleotide and tetranucleotide repeat units found at D11S527 locus

D11S527 locus is recognized by a highly polymorphic DNA probe jg4/pcr, which is a subclone of cosmid 8G11. Cosmid 8G11, also known as c23,23, or cosmid ZD5, was isolated from a human chromosome 11q specific library (Evans, personal communication). The D11S527 locus contains a dinucleotide repeat unit (CA/GT)17 juxtaposed to a tetranucleotide repeat unit (TGTC/ACAG)6. The locus is PCR-based and can be retrieved by primers jg4A 5′-GCCCCTCTACTTGTCTGGAG-3′ and jg4C 5′-ATGCGGCTCCAAGACAAGTTC-3′. It detects 12 alleles in the Caucasian population and has a PIC of .88.

Genomic DNA used for PCR and DNA sequencing

Genomic DNA from 12 unrelated individuals of CEPH reference families were used. These individuals were

known to be homozygotes at D11S527 from the previous genotyping data. They were #1329403, #134102, #140804, #141609, #141811, #1203, #141301, #204, #2304, #4505, #6603 and #88408.

Control DNA sample

The control DNA used, XHM12, is a subclone from a microdissected human chromosome 11q24-q25 specific library. Those microdissected DNA fragments were PCR amplified, digested with Bgl II, partially filled-in and then subcloned into the SalI site of the plasmid vector pTZ18u (microdissected, PCR amplified 11q material was kindly provided by Leland Allen from Dr. Magenis' lab, OHSU). Plasmid DNA from XHM12 was partially purified by using a single colony lysis method (BRL). One half of a single bacteria colony was picked up from the agar plate with a toothpick and transferred into 12 μl single colony lysis (SCL) solution containing 10 mM TrisCL$^-$ pH7.5, 1 mM EDTA, 50 μg/ml proteinase K, and mixed. The lysate was incubated at 55° C for 15 min and then at 80° C for 15 min to denature the proteinase K. The diagram shown in Fig. III-3 illustrates the overall experimental procedures described in this paper.

PCR amplification of D11S527 locus and XHM12

PCR amplification of D11S527 was performed in a total volume of 12.5 μl reaction containing 25 ng genomic

DNA, 5 pmole of primers jg4A and jg4C, 1.5 mM MgCL2, 200 μM dNTPs, 50 mM KCL, 10 mM Tris CL, pH8.3, 0.6 units of Taq polymerase (Perkin-Elmer/ Cetus) and 0.01% gelatin. The DNA samples were aliquoted into 96-well format PCR plates. Amplification was carried out by using the "touchdown" program with denaturation at 94° C for 30 sec, annealing at 66° C for 30 sec, and extension at 72° C for 1 min. Thirty cycles of amplification were performed with annealing temperature decreasing by 1° C for the first 10 cycles (Don et al. 1991). Amplified products were resolved on DNA sequencing gel containing 32% formamide plus 5.6 M urea and transferred onto Hybond N^+ membrane. For detecting alleles, the membrane was hybridized with a 5′-end labeled $(CA)_{15}$ oligonucleotide probe. The PCR primers used for amplifying the DNA insert in XHM12 are XHM13F primer (5′-CAGTGCCAAGCTTGCATG-3′) and IBI reverse primer (5′-CAGGAAACAGCTATGACC-3′). The same PCR condition was applied to amplify the insert of XHM12 as for D11S527, except that we used 2 μl cell lysate and annealing at 50° C for 30 sec.

Single-stranded PCR and dideoxynucleotide sequencing

Single-stranded PCR was performed to generate ssDNA templates for dideoxy DNA sequencing as described by Kaltenboeck et al. with some modifications (Kaltenboeck et al. 1992). One tenth of a total volume of 100 μl

reaction product from standard PCR, as described above, was subjected to the second round linear amplification. Only one of two PCR primers was added into the 2nd reaction; consequently the ssDNA was generated. After single-stranded PCR amplification, the ssDNA template was separated from the PCR primer by centrifugation through CENTRICON 100 column (AMICON). The purified ssDNA was precipitated by centrifugation with 2.5X 0.15 M NaCL/95% ETOH at 4° C for 15 min. In the meantime, one tenth of the centrifuged product was added onto anethidium bromide stained 1.5% agarose gel to make sure that the proper size and amount of ssDNA had been obtained. One third of the final product was subjected to dideoxynucleotide sequencing. By using this combined method, PCR amplified genomic DNAs from two individuals, #1203 and #141301, were sequenced from both strands.

Double strand DNA cycle sequencing

In order to prepare the dsDNA template for cycle sequencing, human genomic DNA was amplified in a total volume of 100 μl by standard PCR, as described above. The dsDNA was then separated from both PCR primers by centrifugation through CENTRICON 100 column (AMICON). After precipitation, the dsDNA was resolved in 14 μl H_2O and 4 μl of the aliquot was subjected to cycle sequencing. DNA cycle sequencing was carried out

according to the manufacturer's instruction (BRL), except the amount of reagents required was reduced by one half.

Results

Allele assessment and molecular basis of polymorphism in (TGTC)n and (CA)n regions at D11S527

Allele assessment for each individual studied was performed by directly counting the number of repeat units from the DNA sequencing ladder and by running PCR amplified alleles with a size standard. Those individuals were chosen both for studying the molecular basis of polymorphism and for studying the slippage at D11S527 locus, by virtue of being homozygotes. Genomic sequencing analysis showed that 9 of those 12 individuals have a fixed number of repeat units in the tetranucleotide region even though they are genotypically different, indicating that the polymorphism seen in those individuals at D11S527 is due solely to the variation in the CA/GT repeat region (Fig. III-4). Three individuals (1/4) have varied numbers of repeat units, not only in the tetranucleotide repeat region, but also in the dinucleotide repeat region. These findings are summarized in Table (III-1).

Table III-1. Allele Assessment and Polymorphism in Tetra- and Dinucleotide Repeat Regions at D11S527 Locus from 12 Unrelated Homozygotes.

indi	allele	No. of repeat in (TGTC)n	No.of repeat* in (CA)n
1329403	A10,A10	6	15
134102	A10,A10	6	15
4505	A10,A10	6	15
6603	A5,A5	6	20
141811	A4,A4	6	21
204	A4,A4	6	21
1203	A4,A4	6	21
141301	A2,A2	6	23
141609	A2,A2	6	23
140804	A5,A5	5	22
2304	A11,A11	7	12
88408	A1,A1	8	19

* These numbers have been reexamined by running PCR products and a ladder of known sequence side by side. Both results were matched very well.

Slippage in D11S527 CA/TG repeat

Dideoxy sequencing method involves DNA synthesis off a single-stranded DNA template by DNA polymerase. The DNA synthesis is initiated by the primer with the complementary sequence to a unique site on the template and is terminated when one of four 2′,3′ dideoxy nucleotide triphosphates is incorporated into the new strands. Those strands are then fractionated on the denaturing gel by electrophoresis. During electrophoresis strands with the same length will comigrate on the gel. In most cases, ambiguous DNA sequences flag possible contamination of DNA templates

caused by untidy preparation or possible mixture of different sized DNA fragments generated by internal deletion and insertion. If there are deletions or insertions within a certain region of these templates, one would expect to see the sequence ladder that is unique 5′ to the deletion or insertion site, but will become ambiguous, even unreadable when one reads pass through the deletion or insertion site. The specific site of the deletion or insertion can be identified precisely by sequencing both DNA strands; this is because the DNA sequence will always be unique at its 5′ region but ambiguous at its 3′ end regardless of which primer being is used. The schematic illustration of sequencing ladders resulting from deletions or insertions in the (CA)n repeat region at theD11S527 locus is shown in Fig. III-5. This is exactly what we observed when we sequenced the entire repeat region from both strands (Fig.III-6). The target sequence, bracketed by primers jg4C and jg4A at D11S527, is shown in Figure (III-7). The entire sequence is 149 bp. By using primer jg4A sequencing from the 3′ end of this region, the tetranucleotide repeat can be read first, followed by the dinucleotide repeat. The sequence from the first nucleotide "G", seen on the gel, to the end of the dinucleotide repeat region, can be read unambiguously either by using double strand DNA cycle sequencing or by using single-stranded dideoxy sequencing method (as shown

in Fig. III-6C). However at the very end of the dinucleotide repeat, bands cross all lanes and the sequence "GAATC" appears several times, strongly indicating that the DNA templates are mixtures of different sized fragments generated by deletion or insertion within this compound repeat region. When the opposite strand, TG strand, was sequenced by using jg4C primer, the dinucleotide repeat was closer to the sequencing primer jg4C. The sequence can be read unambiguously until it reaches the junction of di- and tetranucleotide repeats, where bands in G and C lanes appear at the same position (Fig. III-6B). In the tetranucleotide repeat region, bands in G and C lanes become more or less the same intensity. Since in the tetranucleotide repeat TGTC, T appears in every other position, deletion or insertion of two nucleotides or multiple of two nucleotides within the dinucleotide repeat region would still keep the "T" position matched, even if the deletion or insertion did occur. Consequently, bands appeared at the same position in the G and C lanes. The schematic illustration of such banding pattern resulting from slippage within the poly CA tract is shown in Figure (III-7). Thus, by sequencing directly from both DNA strands within and without the compound repeat region of theD11S527 locus, we were able to conclude that the slippage occurred in the dinucleotide region, not in the tetranucleotide region.

To determine the occurrence of slippage in inserts without any di- and tetranucleotide repeats, XHM12, a subclone from microdissected chromosome 11q24-q25 fragments, was used as a control. It has been shown that XHM12 does not contain (CA)n repeat by colony hybridization with a $(CA)_{15}$ oligo probe and DNA sequencing. The sequencing data obtained from XHM12 showed an unambiguous sequence up to 400 bp from the priming site (Fig.III-6A). Therefore, no deletion or insertion occurred within its 300-400 bp insert.

Discussion

In this study, we have described a simple, powerful method which allows us to directly detect slippage events in a simple sequence tandem repeat region during polymerase chain reaction. Luty et al. (1990) reported the observation of slipped-strand mispairing event at the DXS425 microsatellite marker locus by using an in vivo method. The in vivo method they used has two drawbacks. First, introduction of a (CA/TG)n repeat tract into a bacteria E. coli host generates a number of mutants with a wide spectrum of repeat lengths by slipped-stand mispairing mechanism. The mutation rate is high and increases with longer repeats. Therefore, it is difficult to distinguish the mutation events which occur during DNA replication or the repairing process in

bacteria hosts and those which occur during polymerase chain reaction. In other words, the in vivo method could generate high background. Second, subcloning and screening procedures involved in the in vivo method are labor intensive and time consuming. Compared to the in vivo method, direct DNA sequencing of PCR amplified human genomic DNA has two advantages: 1) only insertion or deletion of repeat units during PCR is considered; and 2) subcloning and screening procedures are avoided, yet produces satisfactory sequencing data.

By utilizing the in vitro method, we have shown that the deletion or insertion event did occur during polymerase chain reaction, and they were most likely generated by slipped-strand mispairing; exhibited by three characteristics, as expected, for SSM. First, deletion or insertion within the (CA/GT)17 dinucleotide repeat region are single unit changes, i.e., they are two nucleotides or multiples of two nucleotides shorter or longer than the original size. In our system, such size changes are most likely due to deletions rather than insertions, since the extra bands seen on the denaturing gel are usually below the most intense allele bands (these bands are considered to be the true allele bands). However, we were unable to confirm directly such deletions from the DNA sequencing data at the D11S527 locus.

Second, we have observed a high frequency of deletion or insertion in the dinucleotide repeat region at the D11S527 locus. In 12 PCR amplified genomic DNA samples that we sequenced, all showed deletions in the dinucleotide repeat region.

Third, slipped-strand mispairing depends on the target sequence. At D11S527, a tetranucleotide repeat (TGTC)6 is adjacent to a dinucleotide repeat (CA/GT)17. Sequencing data of DNA samples from 12 unrelated individuals demonstrated unambiguously that deletion occurred only in the (CA/GT)17 region, not the tetranucleotide region, suggesting strongly that SSM is biased towards dinucleotide repeats. It has been recognized by some investigators that genotyping tetranucleotide repeat polymorphisms is easier than typing dinucleotide repeat polymorphisms because of significantly less extra-band background (Edwards et al. 1991). Since the molecular basis behind it is not fully understood, our finding may shed some light on this issue.

In addition, we have also examined the explanation that a recombinational or "out of register" process may be responsible for generating ladder-like extra bands. Three possible mechanisms of generating truncated DNA molecules have been proposed: 1) brief pausing of the Taq DNA polymerase on the DNA template, 2) premature termination, 3) formation of a partially double stranded

structure of DNA template preventing the synthesis (Meyerhans et al. 1990). Since the premature termination site can occur at any position, the size of recombinant molecules will vary considerably, from one base pair to hundreds of base pairs. More importantly, the deletion or insertion caused by this mechanism should occur in the dinucleotide repeat as well as in the tetranucleotide repeat. The DNA sequences immediately flanking the recombination site should be completely different due to a template-switching mechanism. Although our system is capable of detecting odd number base pair size changes as well as "shuffled" molecules, no such events were observed in all PCR products sequenced. Furthermore, the data from multiplex PCR on D11S527 and D11S534 loci also argues against the annealing of truncated product as the mechanism, because no anomalous PCR products were found in the sizes between D11S527 and D11S534.

It has been demonstrated convincingly that alternating purine-pyrimidine sequences, such as (GC)n, (CA/GT)n and (GA)n, undergo B to Z transition <u>in vivo</u> and <u>in vitro</u> in systems tested (Freund et al. 1989). Such Z-DNA conformation displays a high frequency of spontaneous mutation (range from 1.2×10^{-2} to 3×10^{-4}), and results in deletion and insertion within the dinucleotide repeat region. Deletion is far more common than insertion, and the deletion process stops when the sequence is too short to adopt Z-DNA conformation (about

6 repeat units) (Levinson and Gutman, 1987; Freund et al. 1989). Therefore, the instability of the dinucleotide repeat is probably due to its Z-DNA structure. The biological function of dinucleotide repeat is largely unknown, although there have been some indications of possible functions in transcription and recombination.

In this study, we have also examined the molecular basis of polymorphisms seen at the D11S527 locus. There has been one other similar study so far (Iwasaki et al. 1992). The highly polymorphic feature seen at the D11S527 locus led us to suspect that the variation may occur both in di- and tetranucleotide repeat regions, but we had no idea which repeat would show more variation. The genomic sequencing analysis has shown that the length variation indeed occurs in both regions, with 75% of those variations found absolutely within the dinucleotide repeat region. Although the degree of the variations found in the tetra-repeat is relatively limited—only 3 varied lengths, (TGTC)5, (TGTC)7 and (TGTC)8 have been detected so far—it should triple the variability at D11S527 in 25 per cent of individuals in the population. Despite the small sample size, we believe that the majority of polymorphisms at D11S527 is due to a variable number of repeats within the poly CA tract.

Part IV

The Future Development--Isolation of Polymorphic Markers from the Chromosome 11q Telomere

Introduction

Human telomeres define the chromosome ends. Although there is tremendous diversity in genetic organization and function, the fundamental feature of the eukaryotic telomere appears to be almost completely conserved (Vander Ploeg et al. 1984; Richard and Ausubel 1988; Brown et al. 1990a).

The telomere is subdivided into two regions—terminal region and proterminal region. The terminal region of the human chromosome consists of simple tandemly repeated hexanucleotide TTAGGG. The length of the TTAGGG array varies from 2 to 20 kb, depending on the type of tissues (Cheng et al. 1989). These repeat sequences are highly methylated in the somatic cells. The proterminal region of the human chromosome is also rich in repeated sequences but they are distinct from the terminal simple sequences array. These have been called telomere-associated repeats(see Cheng et. al. for more details). The majority of the repeated sequences from the proterminal region are common to different chromosomes. This has prevented the chromosome

assignment of many telomeric clones obtained so far to a specific chromosome.

Much of our knowledge about the function of telomeres comes from cytological observations. Telomeres protect natural double-stranded DNA ends from degradation, fusion and recombination with chromosome-internal DNA. Telomeres are tethered at the nuclear membrane and are the initial points for chromosome synapsis formation during the zygotene stage in meiosis. It is tempting to interpret this as suggesting the existence of synapsis promoting sequences in the telomere region. However, the function of telomeres at the molecular level is practically unknown.

Two major kinds of approaches have been used to clone selectively telomeric DNA. The first approach is to establish yeast artificial chromosomes (YACs), based on heterologous telomere function in the yeast S. cerevisiae (Cheng, J. et al. 1989). Linear DNA with a centromere, an origin of replication and a single telomere, is unstable as an episome in yeast, but the stability of such a molecule is greatly increased by adding a second telomere. Since the telomere is highly conserved between yeast and humans, it is possible to create stable YACs by adding human telomeres to yeast artificial chromosome halves.

The second approach is to directly clone telomeric fragments into plasmid or bacteriophage vectors (Richard and Ausubel 1988; de Lange et al. 1990). In this strategy, high molecular weight DNA is first digested with Bal31, flush ended and then ligated to the vector sequence. To obtain an enriched telomeric DNA library the DNA fragments could be ligated to the vector sequence to set up two primary endlibraries, in which cloned segments are inserted in opposite orientations. The hybridization of clones from these two primary libraries help to isolate complementary DNA sequences from the telomeric region and discard random clones isolated from broken ends created during the DNA manipulation. The double-stranded DNA molecules are separated from single-stranded DNA molecules before being transformed into a bacteria E. coli host to generate a secondary endlibrary.

Both the direct and the indirect approaches are likely only suitable for isolation of all human chromosome telomeres rather than a particular chromosome end, i.e. the majority of cloned fragments would be rodent in origin when using human-rodent somatic cell hybrids as cloning materials. Among those clones, only 0.2% to 6% will contain telomere sequences (de Lange et al 1990; Cheng et al. 1989). Therefore, the chance of selectively cloning the human chromosome 11q telomere is minimal using these two methods.

In this study, we proposed another method called "protelomeric sequence-interspersed sequence PCR" or "protelomere-INS PCR" and present some very preliminary data. The idea of protelomere-INS PCR is very similar to that of Alu PCR (Nelson et al. 1989). It uses oligonucleotide primers derived from the human protelomeric region and Alu or LINE sequences to amplify specifically DNA sequences between the protelomere region and nearby Alu or LINE sequences. Those amplified protelomeric sequences will be subcloned into appropriate vectors, e.g., plasmid vector pTZ18u, and then detected for RFLP, VNTR or (CA/GT)n microsatellite polymorphisms. DNAs from human-rodent somatic cell lines, containing the intact chromosome 11 or partial long arm of the chromosome 11 as the only human material, will be used as templates for PCR. Since the DNA sequences from the human telomeric region is different from that of the rodent, only the human DNA can be amplified (de Lange et al. 1990). The identification of telomere sequences contained in YACs, plasmid or bacteriophage clones can be accomplished by one or more of three methods: 1) hybridization with previously mapped telomeric or protelomeric DNA probes, 2) detection of Bal 31 sensitivity of the DNA sequence, 3) <u>in situ</u> hybridization using telomeric specific DNA probes (Fig. IV-1).

Materials and Methods

Cell lines and control DNAs

DNAs from human-rodent somatic cell lines MC-1, J1-11, J1-44 and TGD5D1-1 were used as templates for polymerase chain reaction and have been described in Part I. A mouse cell line (RAG) and a hamster cell line (E36) were used as control DNAs (Litt et al. 1986).

Amplification from somatic cell hybrid lines and parental cell lines

Alu primers were designed according to the literature with some modifications (Nelson et al 1989; Ledbetter et al. 1990; Cotter et al. 1990). At the 5' end of primers v599pst (5' AAGTCTGCAGTTGCAGTGAGCCGAGAT 3') and v517 (5' GGACGATCATCTCAGCTCACTGCAA 3'), a PstI cleavage site and a Sau3A site were added on, respectively. An oligonucleotide primer vpTH2 was designed from published human protelomeric sequence data (de Lange et al. 1990). It was derived from the 5' end of a 390 bp CpG rich region, which is next to the telomeric repeats TTAGGG. The total length of the CpG-rich region and the telomere repeats was about 700 bp and has been cloned by de Lange et al. (1990). This subclone, also referred to as pTH2-delta, hybridized strongly to human chromosome 7, 16, 17 and 21q, and hybridized moderately to human chromosome 3, 4, 5 and X chromosome. Since none of the somatic cell lines used by de Lange et al. contains chromosome 11 as the only

human material, it is not clear whether this probe would also hybridize to human chromosome 11. Subclone pTH2-delta was chosen for designing a PCR primer because it did not cross-hybridize efficiently with rodent DNA (de Lange et al. 1990). vpTH2 and its derivative vpTH2EcoRI have the same sequence except at the 5′ end, where vpTH2 carries a Sau3A cleavage site, whereas vpTH2EcoRI carries an EcoRI site (vpTH2: 5′ GTGATCGTTACACGCGAACAGTCTGA 3′; vpTH2EcoRI: 5′ GTGAATTCGTTACACGCGAACAGTCTGA 3′). PCR was performed in a total volume of 25 μl containing 100 ng DNAs from somatic cell hybrid lines and rodent parental lines, 2.5 pmole of each primer, 1.5 mM MgCl2, 200 μM dNTPs, 50 mM KCL, 5 mM TrisCl⁻, pH8.3, 0.3 unit of Taq polymerase (Perkin-Elmer/Cetus) and 0.01% gelatin. Thirty cycles were performed for 1 min at 94° C (denaturing), 1 min at 56° C (annealing) and 5 min at 74° C (extension). The second-round PCR was carried out using the same conditions as described above, except that 5 μl of product from the first round PCR was used as the template and the annealing was at 62° C. Amplified DNA fragments were analyzed by running on the ethidium bromide stained 1-2% agarose gel.

Cloning of protelomere-INS PCR products from somatic hybrids

Amplified PCR products were separated on the low melting agarose gel. The bands not present in amplified rodent DNAs were sliced out of the gel and purified through CENTRICON 100 column, following the instruction from supplier (AMICON). The purified DNA was digested with either Sau3A or EcoRI plus PstI and was then cloned into SalI (Sau3A-end fragment) or EcoRI-PstI sites (EcoRI-PstI-end fragment). Ligation was performed in a total volume of 20 μl with 10x ligation buffer and T4 ligase, following the instruction from supplier (BRL).

Direct sequence analysis

DNA sequencing was performed by the using Sequenase kit following the instruction from supplier (USB).

Results

Two major kinds of background in our system must be assessed in order to ascertain that appropriate PCR products have been obtained. One is non-specific amplification of inter-Alu sequences by a single Alu primer; another is non-specific amplification of rodent DNAs from somatic cell hybrid lines. The non-specific amplification primed by one Alu primer was assessed by a number of trials including PCR, using one of two Alu primers; PCR amplification, using only the protelomeric primer; and PCR, using an Alu primer and the protelomeric primer. Likewise, the degree of amplification of rodent

DNA was also detected. As shown in Fig. IV-2A, when using only one Alu primer in the PCR at lower annealing temperature, various sizes and amounts of DNA were generated from somatic hybrids as well as RAG (mouse) and E36 (hamster). However, by increasing the stringency of PCR conditions, e.g. annealing at 62° C instead of 56° C, such non-specific amplification was decreased substantially (shown in Fig. IV-2D, right panel). In contrast, when using only the protelomeric primer, no amplified product was found in all cell lines used (data not shown). When using an Alu primer and the protelomeric primer, various amounts of amplification from all 4 somatic hybrids, as well as the hamster cell line, were observed. There was very little amplification detected from the mouse cell line (Fig. IV-2C, IV-2D left panel).

PCR products from J1-44 and TGD5D were further amplified by the second-round PCR. After being separated on the low melting agarose gel, DNA fragments of interest were purified and subcloned into pTZ18u (see Materials and Methods). Eight white colonies were randomly picked up. By electrophoresis on the 2% gel, it was shown that the size of inserts varied from 300 bp to 4 kb (Fig. IV-2B).

It was important to establish whether these subclones have inserts that are from PCR products and from the protelomeric region of human chromosome 11. Two

of them, TGD5D/1 and J1-44/1, were sequenced by using M13 universal and reverse primers. TGD5D/1 is a subclone isolated from somatic hybrid line TGD5D and contains a 300 to 400 bp insert. Sequencing analysis of TGD5D/1 showed sequences of v559pst and vpTH2EcoRI primers at the vector/insert boundaries, indicating that it has a PCR amplified fragment. However, analysis of the entire sequence of this subclone revealed that its insert is actually an inter-Alu fragment, because we found Alu sequences not too far from its ends when compared with the revised Alu consensus sequence (Kariya et al. 1987). J1-44/1 is a subclone isolated from somatic hybrid J1-44 and contains a 300 to 400 bp insert, which was originally a 700 bp fragment prior to the subcloning. Sequence analysis of J1-44 has failed because it contained heterogenous DNA templates.

Discussion

The telomere region on the long arm of human chromosome 11 is the last gap on the 11q index map, and thus encouraged us to test the possibility of cloning this region selectively. Although the method used in this study did not lead to success, better understanding of pitfalls that may have caused the failure will help us to design more efficient methods for isolating the polymorphic markers from the protelomere region of chromosome 11q in the future. First, the protelomeric

primer is not specific to the protelomeric region. Consequently, the inter-Alu region was amplified at least in one subclone sequenced. Second, the second-round PCR had both positive and negative effects. It enriched PCR amplified products from the 1st PCR, making the subcloning easier. The down side of this procedure is that small fragments tend to amplify much more efficiently than larger ones. Consequently they become over- represented in the final product, and longer fragments are most likely missing in the later cloning process. Third, although it has been shown that the length of the human protelomere region varies from 4 kb to 100 kb (Cheng et al. 1989; de Lange et al. 1990), PCR failure could result if the protelomere region on the 11q is about 10 kb long. This occurs because the Taq polymerase used is not capable of amplifying fragments within this range routinely.

In retrospect, we should investigate the possibility of using more direct and efficient ways to achieve our goal, e.g. using a (TTAGGG)n oligonucleotide probe to screen preexisting human chromosome 11 or chromosome 11q specific libraries (Nakamura; Evans, pers.comm.). From such screening the positive clone may contain the chromosome ends. In fact, there have been a few studies which demonstrated successful cloning of the telomere region of other chromosomes by utilizing this method (de Lange et al. 1990; Ijdo et al. 1991). Several groups

around the country are actively involved in cloning telomeres of human chromosomes using YACs (Thomas Shows, personal communication). Although the chromosome 11q telomere has not yet been found among those YAC clones, continued efforts will probably lead to this finding in the near future.

Conclusions

Eight highly informative simple sequence repeat polymorphisms have been isolated and mapped to specific bands on the long arm of human chromosome 11 by using somatic cell hybrid panels, in situ hybridization and linkage analysis. Five of them are index quality markers with heterozygosity or PIC >0.70. All of them can be genotyped by PCR, and two of them (D11S527/S534) have been typed simultaneously by multiplex PCR. An index map of human chromosome 11q has been constructed, mainly based on index markers developed by us and by two other groups. The index map is anchored by marker loci D11S288/S436 on the short arm at the 11p11 region and ends at D11S439. It covers the entire long arm of chromosome 11 except the very distal region (11q25-qter). The sex-averaged index map of the chromosome 11q spans 82 cM, whereas the sex-specific index maps span 118 cM and 53.4 cM for female and male, respectively. The genetic distance between the two adjacent index markers varies from 1.4 cM to 14.7 cM. The order of the index markers is cen-
S228/S436-sMSH3-sHI1-D11S97-INT2-D11S534-D11S533/S527-D11S388-sMSH2-D11S35-Mfd231-DRD2-D11S144/S351-D11S490-6d11-D11S439-qter. These highly informative markers and the well-constructed linkage map will undoubtedly be useful

for mapping unknown disease genes, and for positional cloning of disease genes localized grossly on the 11q, including multiple endocrine neoplasia type I, ataxia telangiectasia, tuberous sclerosis type II, etc. Both possible germline and somatic mutations have been found at two index marker loci (D11S534/S439). The Mendelian inheritance of one of the three new mutations was observed. We see no bias towards longer and more polymorphic microsatellite markers. The nature of all mutations are either gain or loss of two nucleotides. Since this feature corresponds very well with one of the features of a slipped-strand mispairing event, we believe that the new mutations described here were generated by the slipped-strand mispairing mechanism.

In this study, we have described a simple, powerful method which allows us to directly detect slippage events in a simple sequence tandem repeat regions during the polymerase chain reaction. All three major characteristics expected for slippage have been observed in the dinucleotide repeat region examined. These features include: 1) deletions or insertions that are integers of one unit changes, although deletions are far more common than insertions; 2) deletion or insertion occurs at very high frequency, virtually in every DNA sample examined; 3) slippage depends on the target sequence. It is biased towards dinucleotide other than tetranucleotide repeats. We were able to exclude the

possible explanations that recombination and/or "out of register" events during PCR would generate the anomalous products. Our final conclusion is that the anomalous products are generated by slippage. The level of the length polymorphisms between dinucleotide and tetranucleotide repeats are very different. Seventy five percent of polymorphisms seen at D11S527 which we have examined are due to variations in the dinucleotide region. Only 25% of polymorphisms are due to variations in both dinucleotide and tetranucleotide regions. This observation, again, supports the hypothesis that slippage may play a major role in the repetitive DNA sequence evolution by producing a larger number of polymorphisms.

The telomere region of chromosome 11q remains as the final gap region on the current index map. We have proposed the method called protelomere-INS PCR to selectively clone this region. Although we did not succeed in the first try, combined methods and efforts with other groups would be useful in the future.

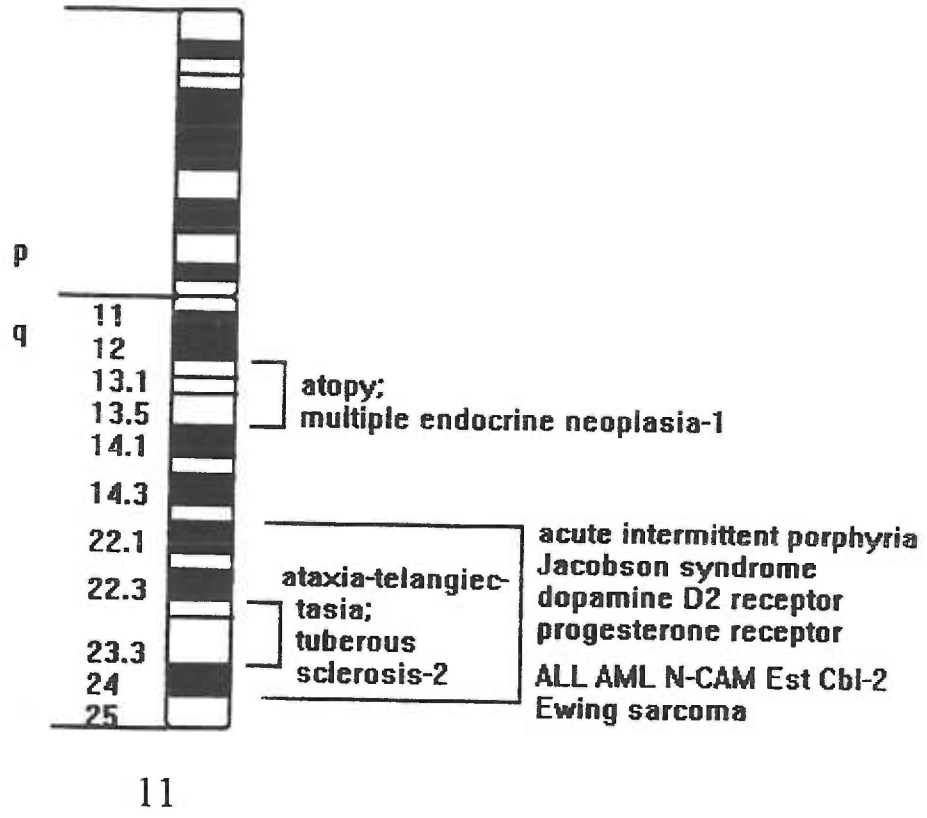
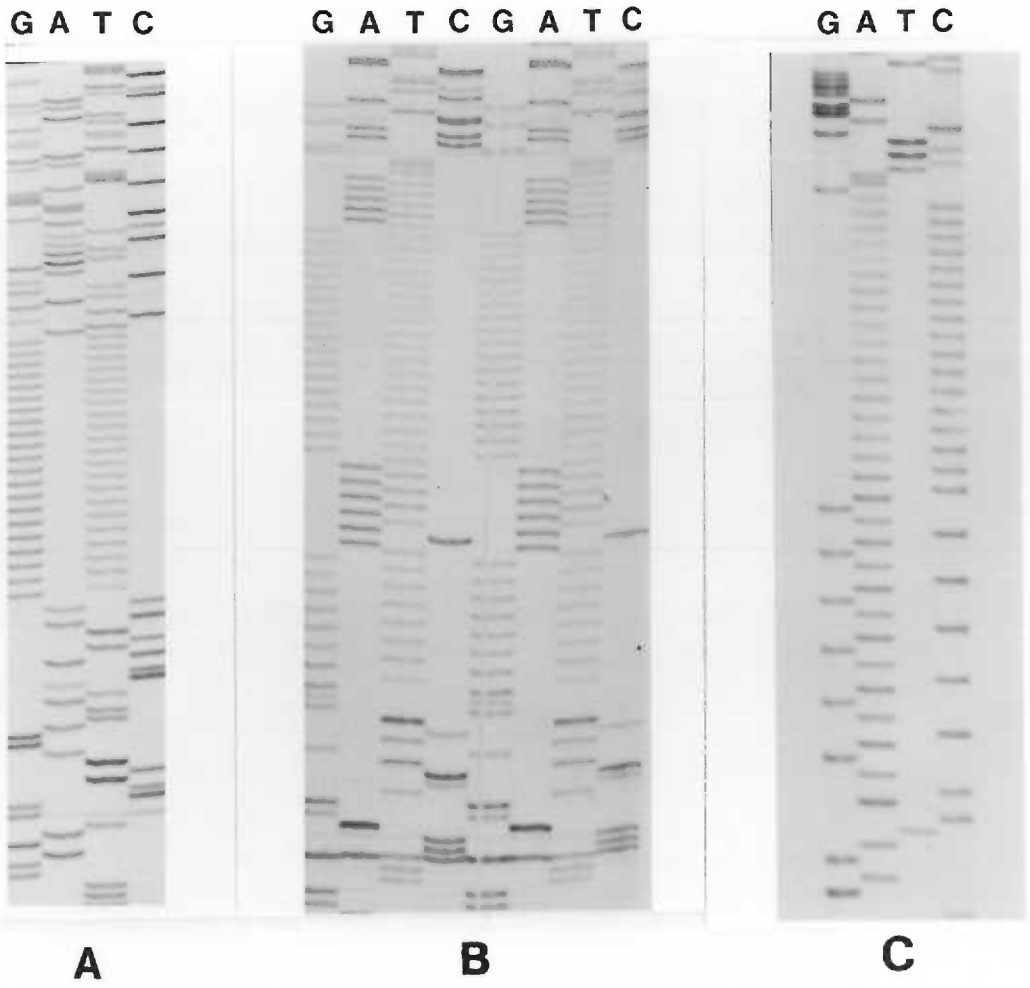


Fig. I-1. Diagram of the long arm of human chromosome 11, with locations of genes/disease related genes.

Fig.I-2. Three classes of microsatellite repeats. A. Perfect repeat; a (GT)21 repeat found in cosmid 9/4. B. Interrupted repeat; a (GT)8(AT)5(GT)14 repeat found in cosmid 1/16. C. Compound perfect repeat; a (ACAG)6(AC)17 repeat found in cosmid 8G11.


G A T C
G A T C G A T C
G A T C
A
B
C

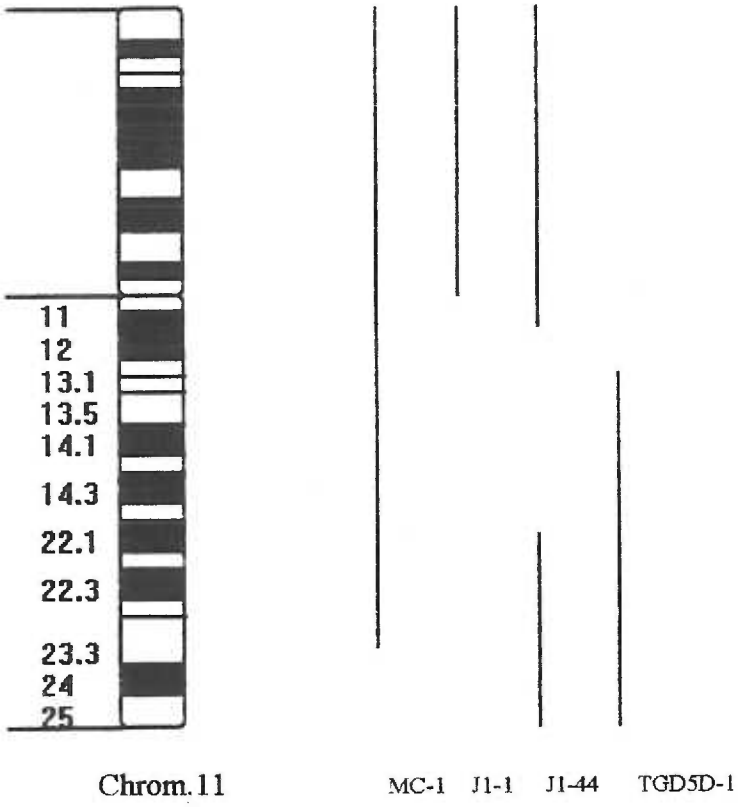


Fig. I-3 Human chromosome 11 regions present in 4 somatic cell hybrid lines.

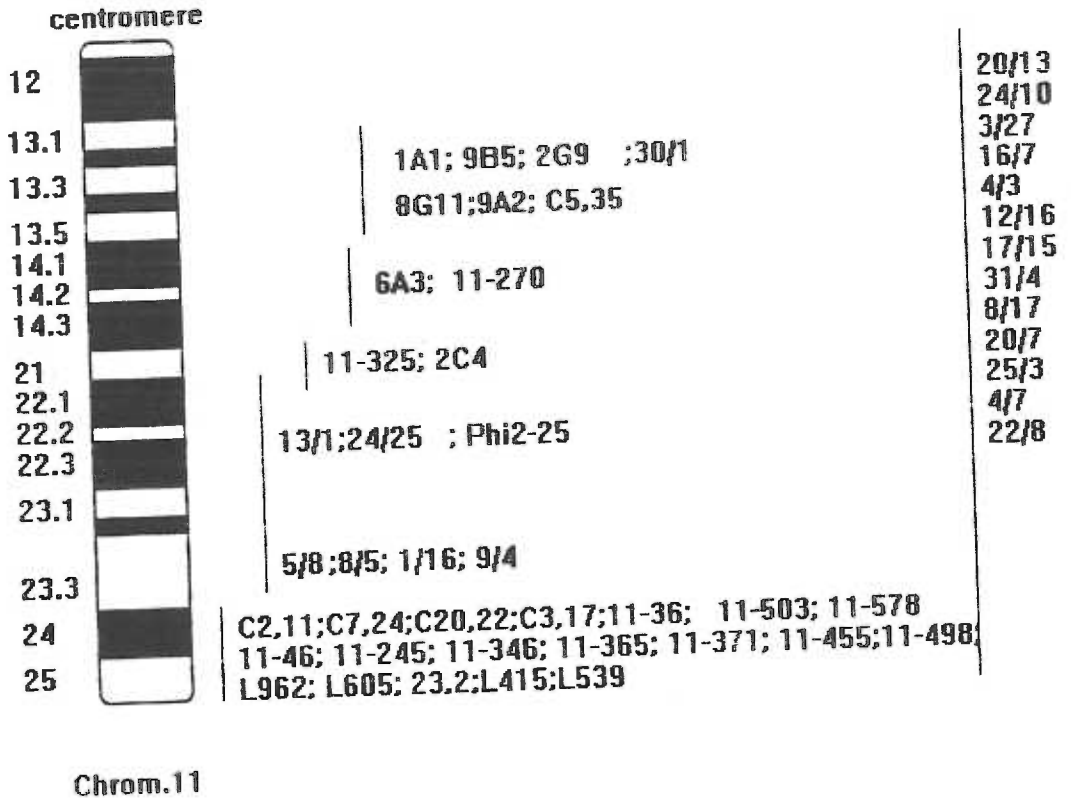


Fig. I-4. Illustration of the chromosomal locations of fifty cosmids isolated from the long arm of human chromosome 11.

Fig I-5A. Genotyping microsatellite marker D11S534 in a CEPH family by using internal labeling method. Each allele shows 2 discrete bands. The faster moving band represents the CA strand whereas the slower moving band represents the GT strand.

Fig I-5B. The conservation between human and hamster in one of non-polymorphic dinucleotide repeats 2/11. The random individuals from CEPH, hamster (E36), mouse (RAG), cosmid 2/11 and their corresponding alleles from PCR amplification of the 2/11 dinucleotide repeat are shown. Note that the hamster shares the same allele as humans, whereas the mouse exhibits 2 allelic polymorphism (arrow).

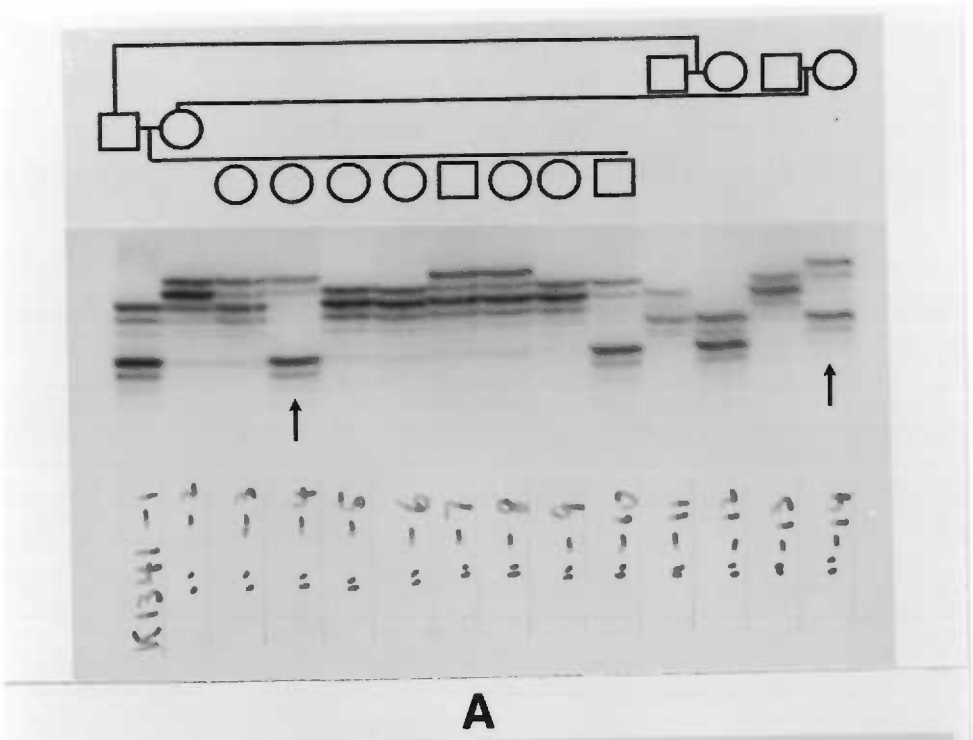
K1345
1345-1
1345-2
" -2
1345-3
" -4
" -5
" -6
" -7
" -8
" -9
" -10
" -11
" -12
" -13

A

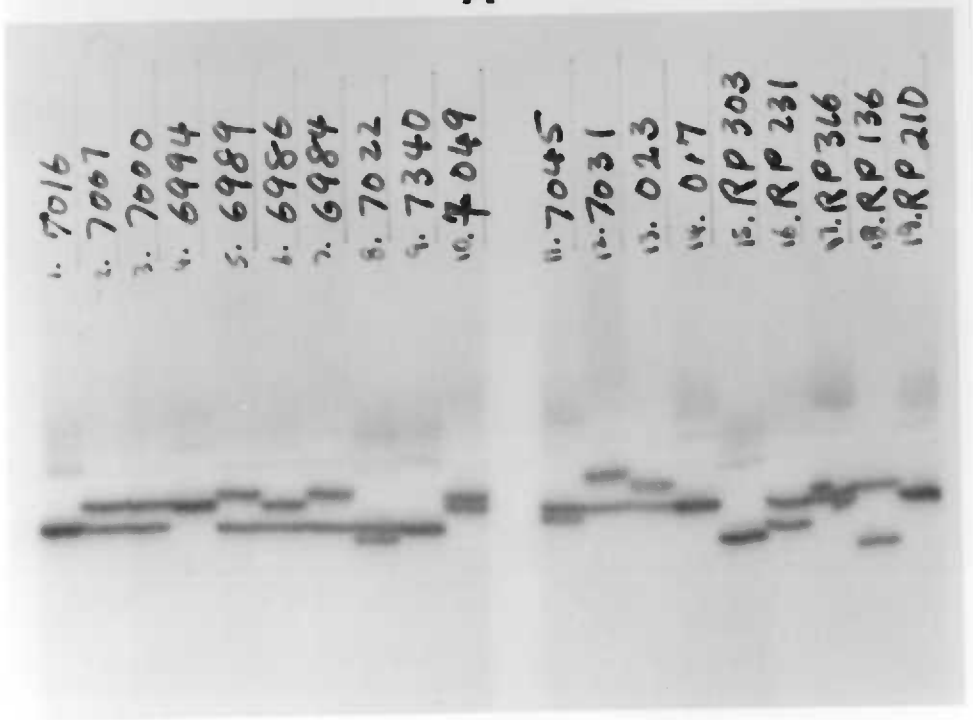
133401
133402
134001
134002
134101
134102
134401
134402
134501
134502
E36
RAG
cosmid 2,11

B

Fig.I-6. Genotyping the microsatellite marker D11S534 in the CEPH family K1341 (A) and 19 unrelated individuals (B) by using end labeling method. Note that each allele shows a strong main band as well as 1 to 3 "shadow" bands (arrows in A) which may interfere with scoring. However, by using blotting method, those "shadow" bands have been successfully eliminated (B).


K1341-1
1:-2
1:-3
1:-4
1:-5
1:-6
1:-7
1:-8
1:-9
1:-10
1:-11
1:-12
1:-13
1:-14

A


1. 7016
2. 7001
3. 7000
4. 6994
5. 6989
6. 6986
7. 6984
8. 7022
9. 7340
10. 7049
11. 7045
12. 7031
13. 023
14. 017
15. RP303
16. RP231
17. RP346
18. RP136
19. RP210

B

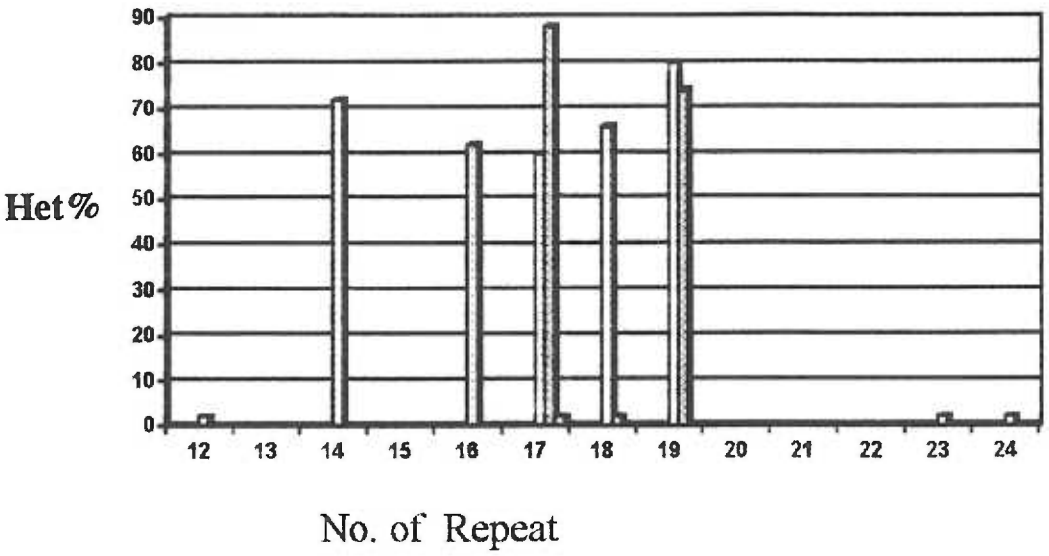


Fig. I-7. The number of repeat unit vs. the percent of heterozygosity of twelve simple-sequence tandem repeats isolated from human chromosome 11q cosmids.

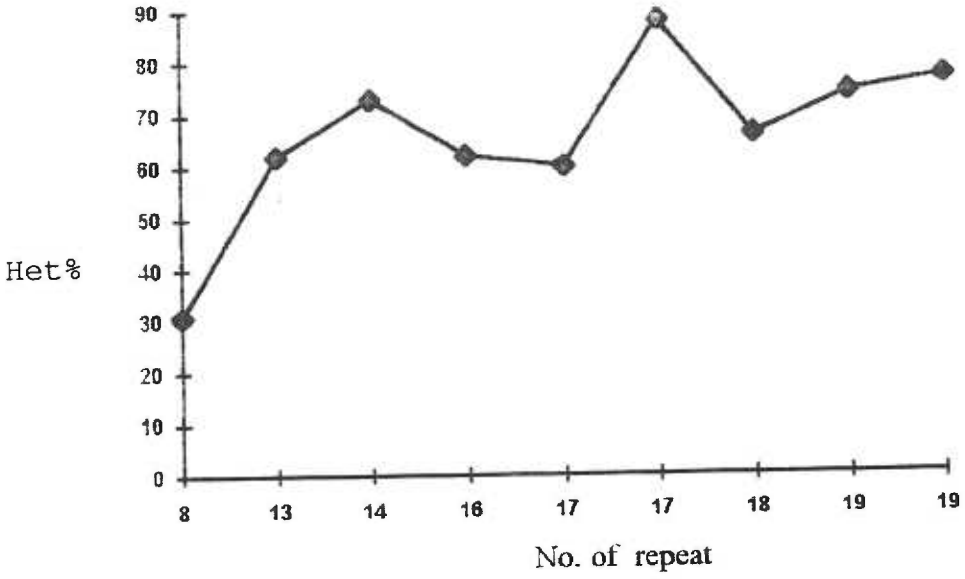


Fig. I-8. Diagram of the relationship between the number of repeat and the percent of heterozygosity of nine microsatellite markers from chromosome 11q (D11S83, c535, D11S490, DRD2, D11S528, D11S527, D11S420, D11S534 and D11S439, from the left to the right).

Fig.I-9. Genotyping two microsatellite markers (D11S527 and D11S534) by using multiplex PCR. A. PCR products were loaded twice with a 20 min interval between loading, and resolved on a denaturing gel without formamide. The two upper groups of bands represent the D11S534 locus and the lower groups represent the D11S527 locus. The membrane was hybridized with 5'-end labeled (CA)15 oligonucleotide probe. Note that a few individuals (see arrows) have questionable genotypes. B. The lower portion of a gel showing the PCR products from the D11S527 locus. The same PCR products and electrophoresis condition were used as for A, except that we used a formamide-urea gel. Note that the background has been substantially reduced, making scoring easier.

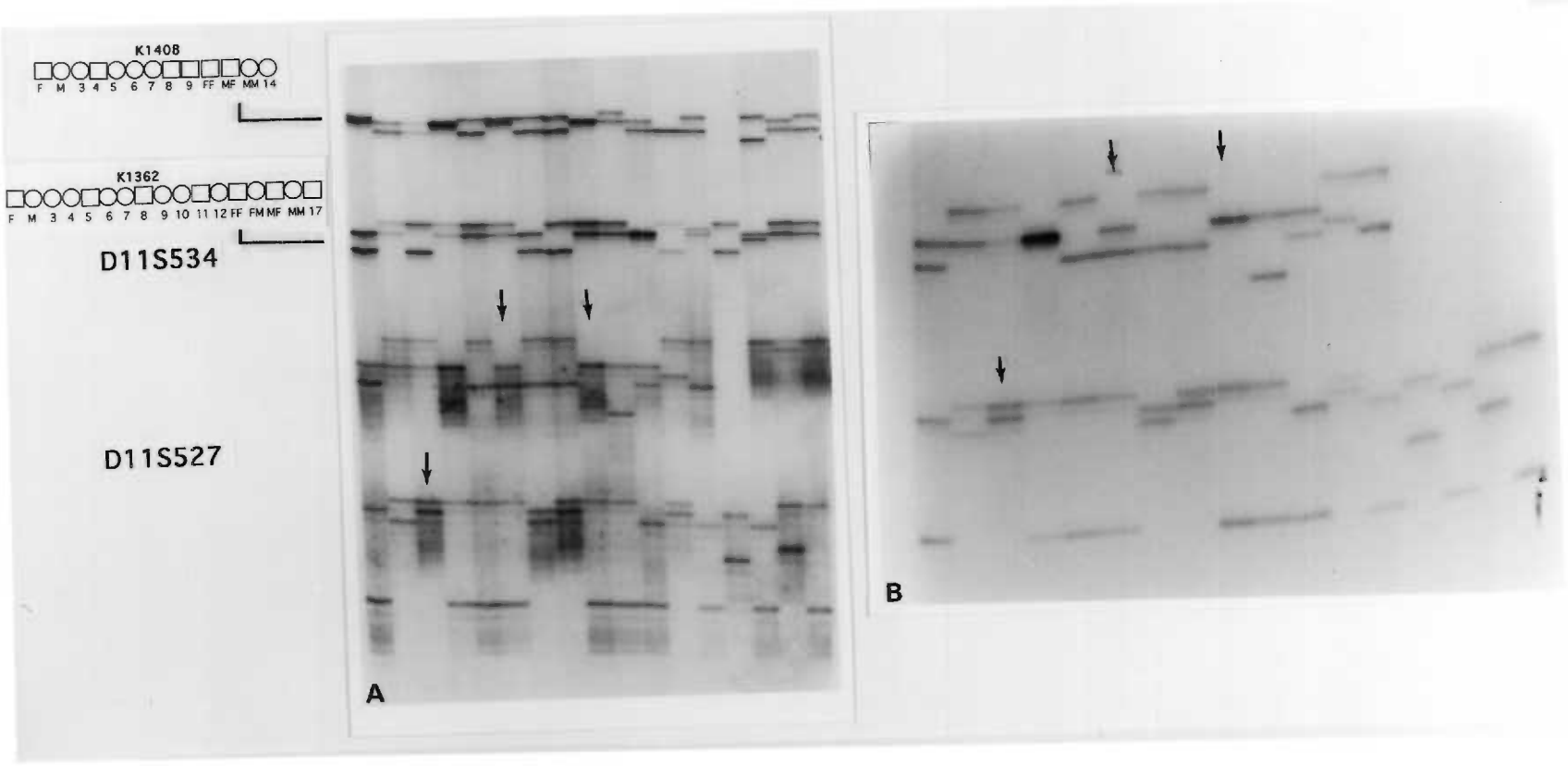
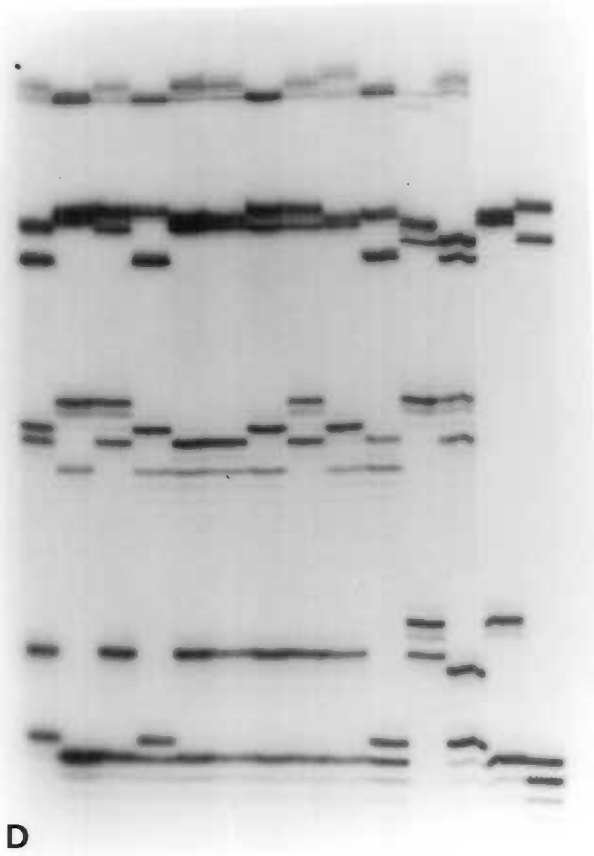

K1408
F M 3 4 5 6 7 8 9 FF MF MM 14
K1362
F M 3 4 5 6 7 8 9 10 11 12 FF FM MF MM 17
D11S534
D11S527
A
B

Fig.I-9. C. The same membrane shown in Fig.I-9A was hybridized with a 5'-end labeled locus specific oligo (LSO) specific to the D11S534 locus. D. Formamide-urea gel showing typing of D11S527 and D11S534 in CEPH families K1375 (the 1st and 3rd group of bands from the top) and K1341 (the 2nd and 4th group of bands). The membrane was hybridized with 5'-end labeled (CA)15 ligonucleotide probe.

D11S534

C



D11S534

D11S527

D

Fig.I-10. Direct cosmid DNA sequencing of cCI11-415 by using cycle sequencing method. Sequencing was performed by using the whole cosmid DNA as template and initiated by oligonucleotide primer XH415-1. Sequence begins at "G" which corresponds to base number 3 in Fig.I-12 and continues to beyond 100 bp on the top (bars in both right and leftthe panels). Sequencing products were loaded twice on the 6% denaturing gel with a 1.5 hr interval between loadings as shown in the left, the middle (the 1st loading) and the right (the 2nd loading) panels. Sequence ladders shown in the left and the middle panels are same, except that the left ladder is more intense.


G A T C
G A T C
G A T C

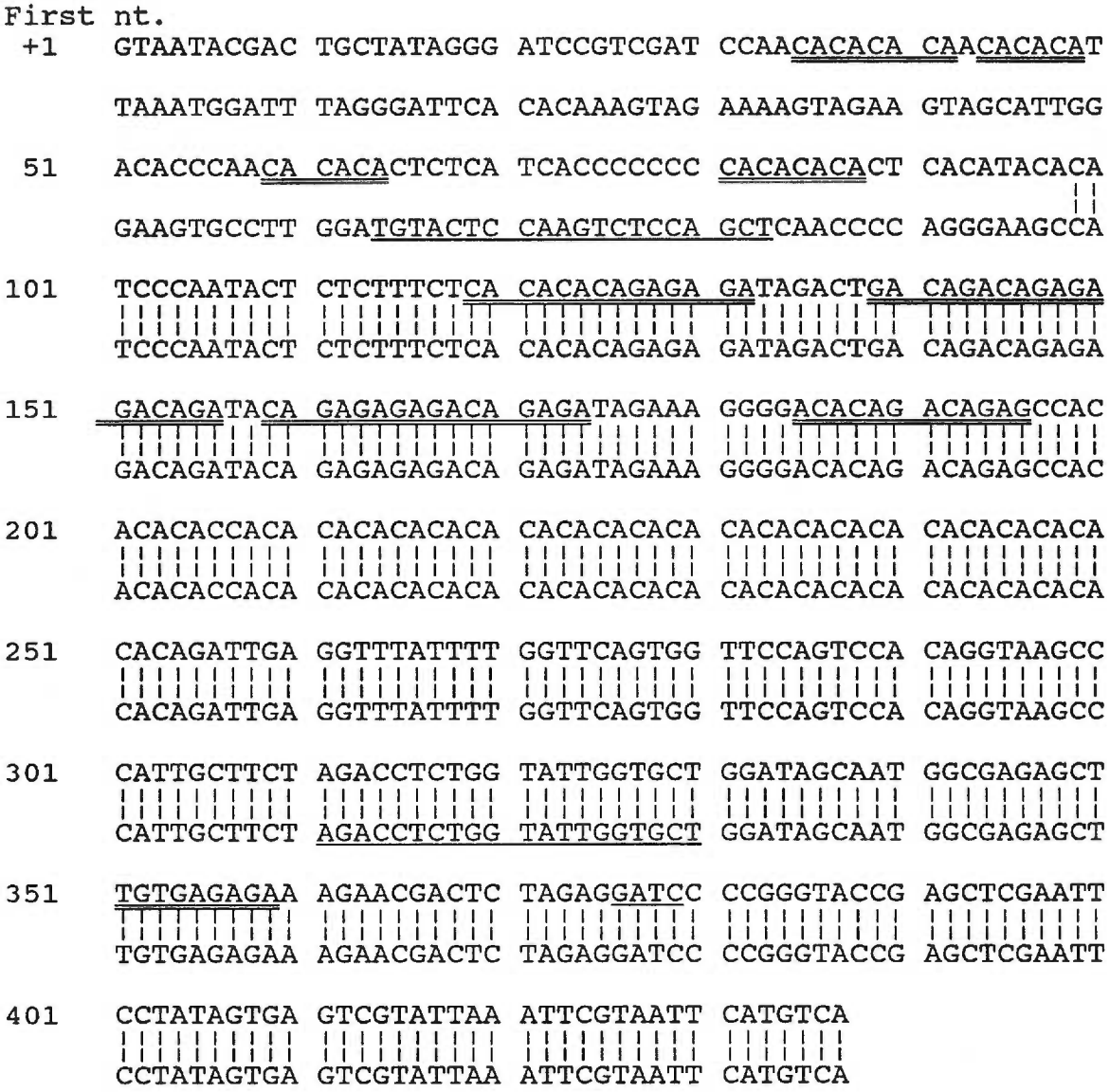


Fig.I-11. The 437 bp long DNA sequence from the cosmid cCI11-415 (upper line) and the plasmid XH415 (lower line) which flanks the recombination site in the plasmid XH415. The primer sequences for both XH415-1 (from nucleotide No.64-83) and XH415-2 (from nucleotide No.311 to 330) are underlined and the homologous sequence between cosmid cCI11-415 and plasmid XH415 is marked by vertical lines. The short simple sequence tandem repeats are marked by double underlines.

```
  +1  GAGAGTATTG GGATGTGTAT GTGAGTGTGT GTGGGGGGGG TGTATGAGAG
      ******       ******* ***  ***** ********** **********
 +51  TGTGTGTTGG GTGTATGTGT GTTGTGTGTG TTGGATCGAG GGATCCCTAT
      ********** ********** ** ******* *                  ***
+100  AGTCAGTCGT ATTATGCCGC CGCGAATTCT CATG
      *        * ********** *   ****
```

Fig.I-12. DNA sequence of the cosmid cCI11-415 extended from the primer XH415-1. The corresponding position of the sequencing primer XH415-1 in the plasmid XH415 DNA is shown in Fig.I-11 (Primer XH415-1 starts at nucleotide No.64). Short simple sequence tandem repeats are marked by asterisks.

```
  +1  TCTAGGATTA TACCACTCTA CACGTGTGTG TGTGTGTGTG TGTGTGTG
                               ******* ********** ********
 +51  TGTGTGTGTG TGTGTATGCT GATGTGCAGA TATAC.
      ********** *****
```

Fig.I-13. DNA sequence flanking the $(GT)_{21}$ repeat in the plasmid 9/4-1. The PCR primer sequences are underlined; the repeat sequence is marked by asterisks.

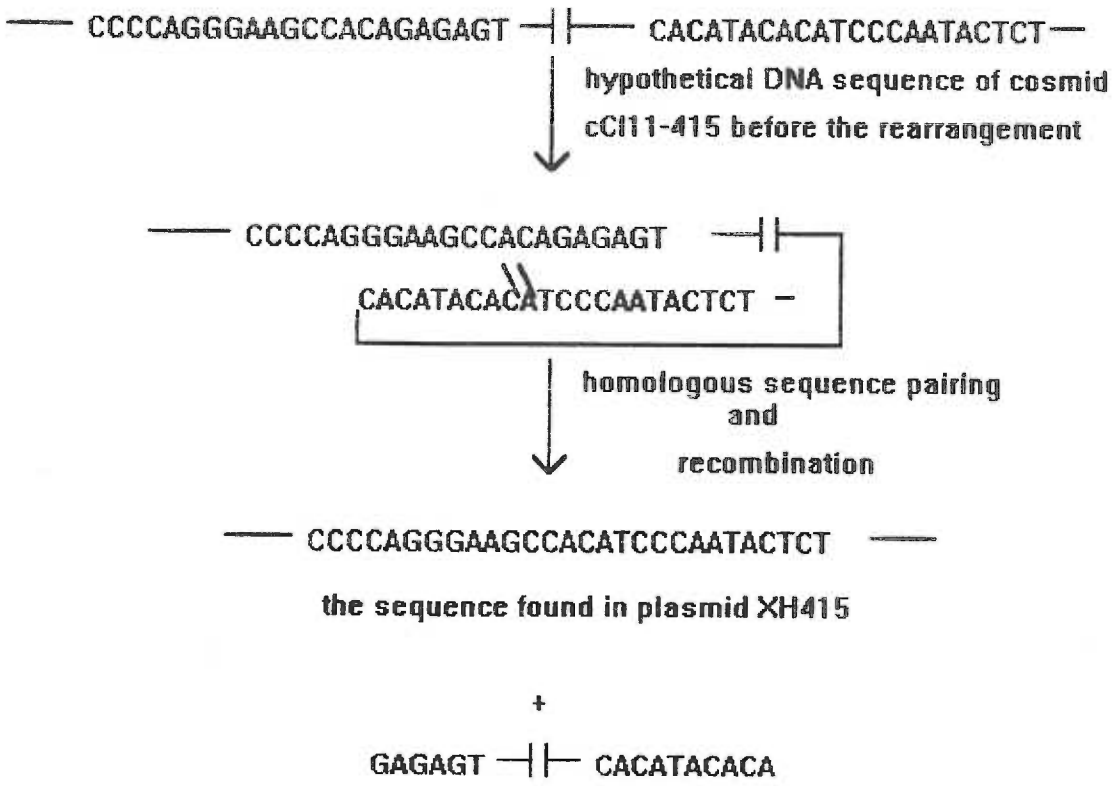
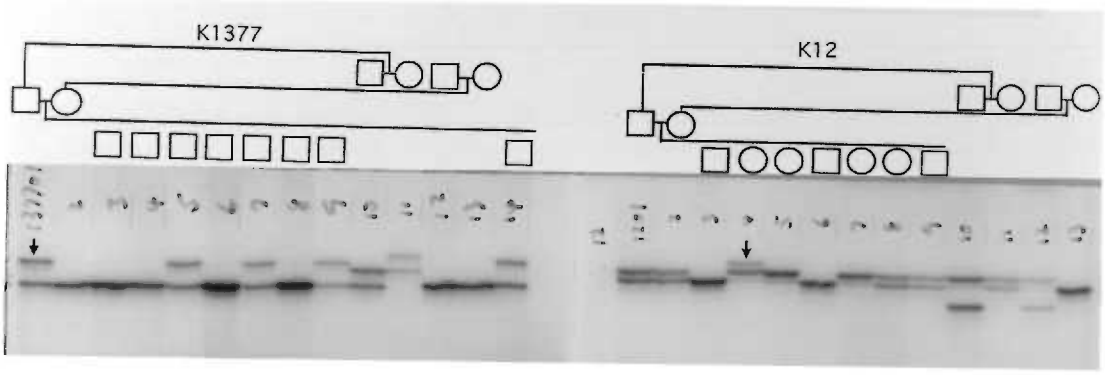


Fig.I-14. Hypothetical recombination which might have occurred in the cosmid cCI11-415.

Fig.I-15. Spontaneous new mutations found at microsatellite marker locus D11S534 in 2 CEPH families (K1377,left; K12, right). Individual 137701 inherited one of alleles (A6) from his father, but his another allele (A3; arrow) is inconsistent with either alleles of his mother. Note that the new mutation has been transmitted to his offspring, indicating a possible germ line mutation. Likewise, one of alleles (arrow) of individual 1204 is inconsistent with both her parents.


K1377
K12

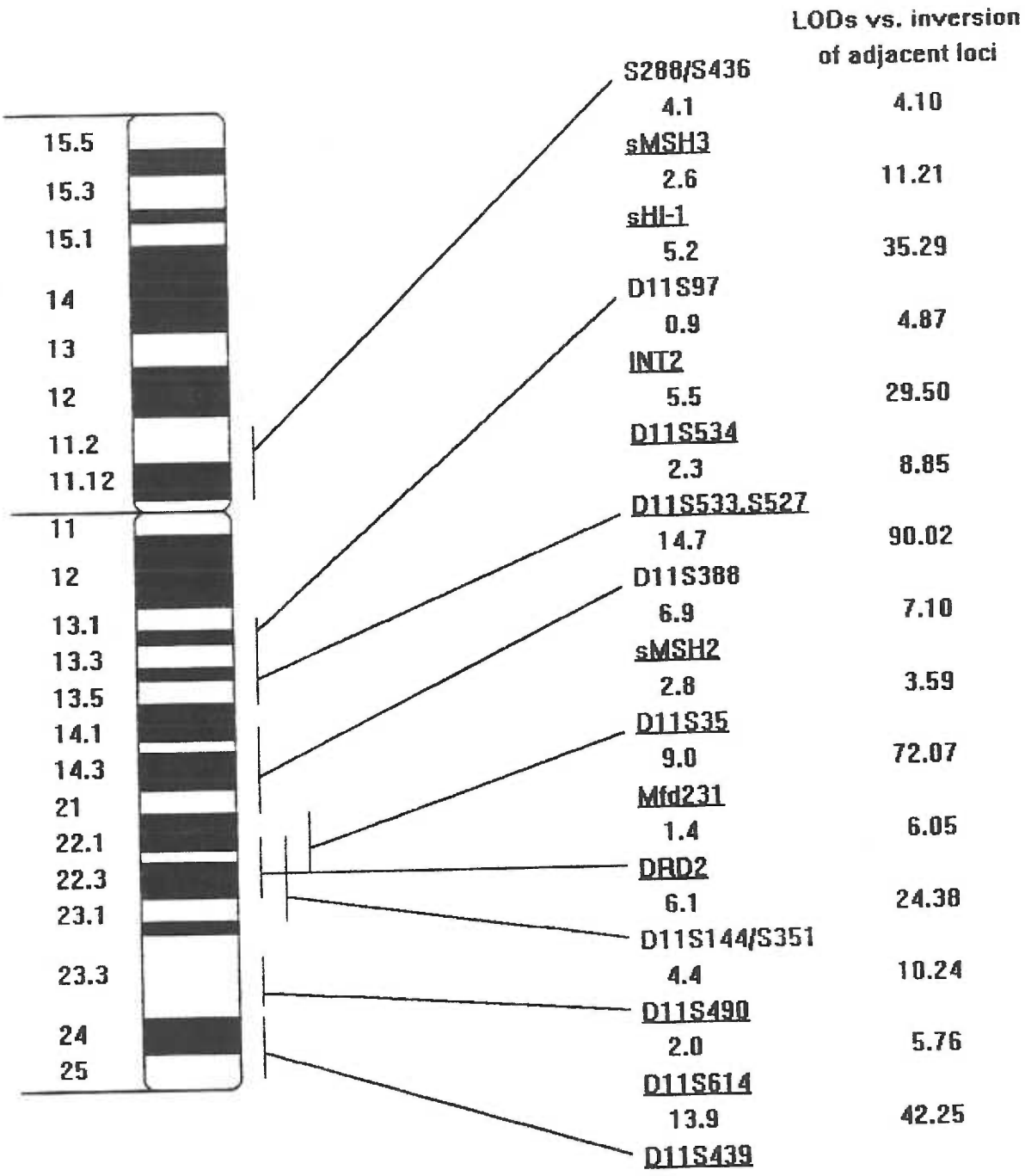


Fig.I-16. Sex-averaged index map of the long arm of human chromosome 11. Seventeen markers with index marker quality were included. The bars show the chromosomal locations of markers. The map is scaled in Kosambi centimorgan. The log10 likelihood of odds against the inversion of adjacent loci are shown on the right.

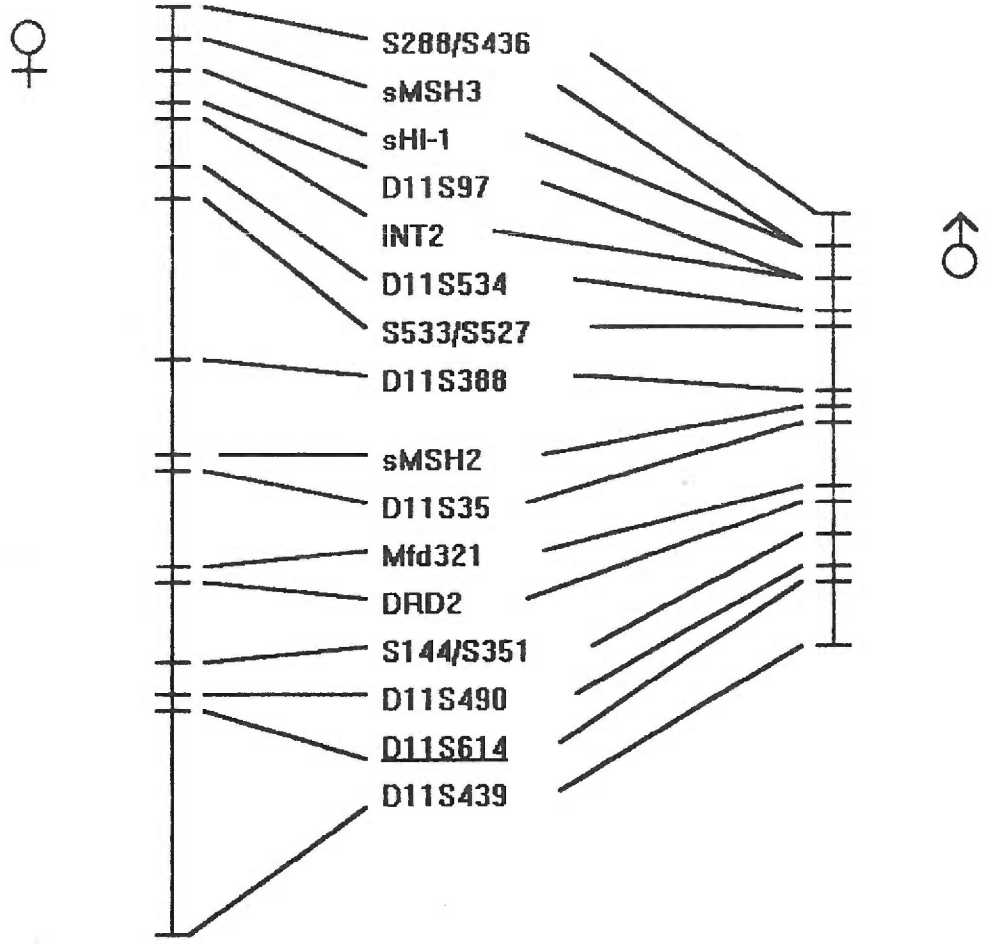


Fig.I-17. Sex-specific index maps of the long arm of human chromosome 11. Seventeen markers with index marker quality are included. The map is scaled in Kosambi centimorgan.

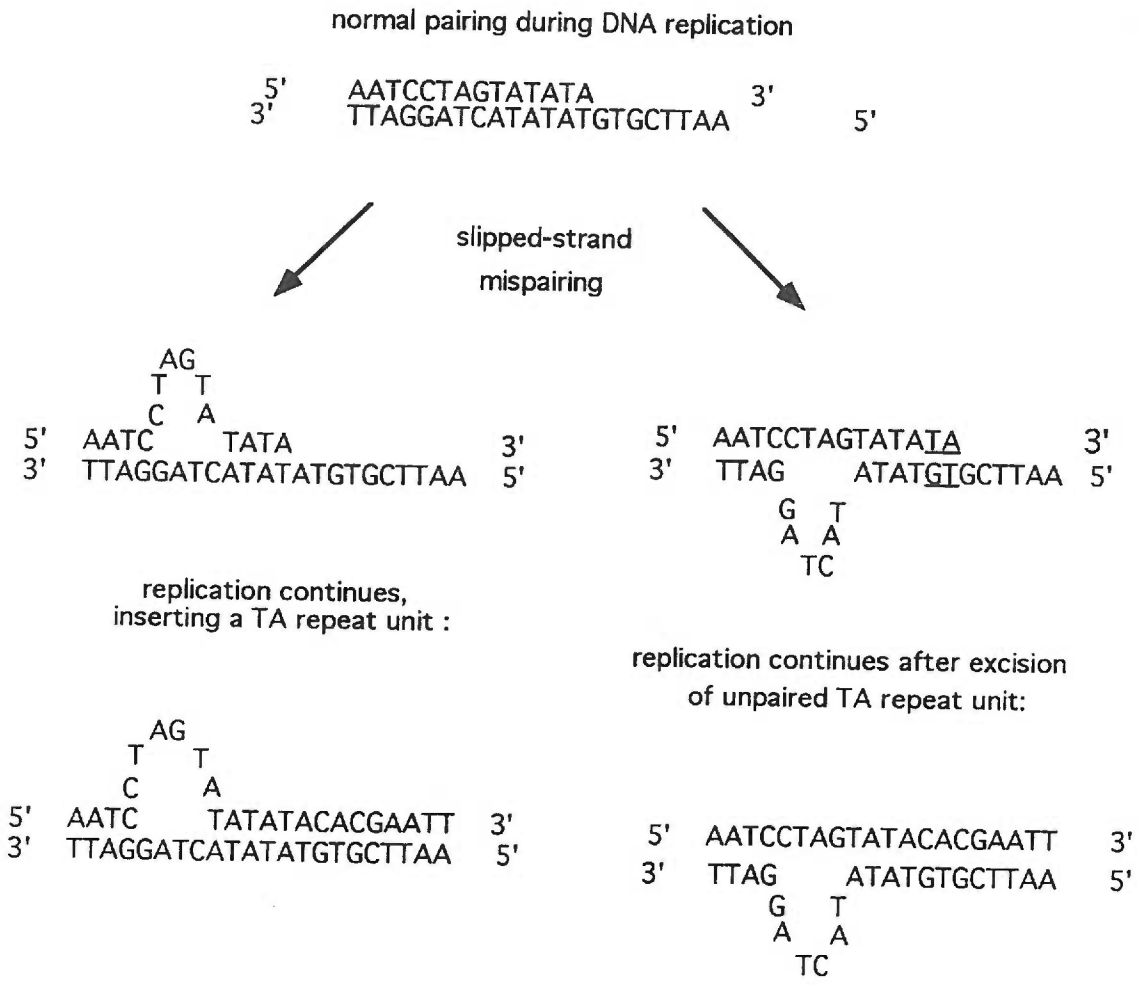


Fig.III-1. Generation of duplications or deletions by SSM between contiguous repeats. 2-Base slippage in a TA-repeat during replication of a DNA dulpex, followed by continued chain elongation. Slippage in the 3'-5' direction (left panel) results in insertion of one TA unit; slippage in the other direction (right panel) results in deletion of one repeat unit. Unpaired repeat units are removed presumably by DNA polymerase (underlined) (Courtesy of Levinson and Gutman, 1987b).

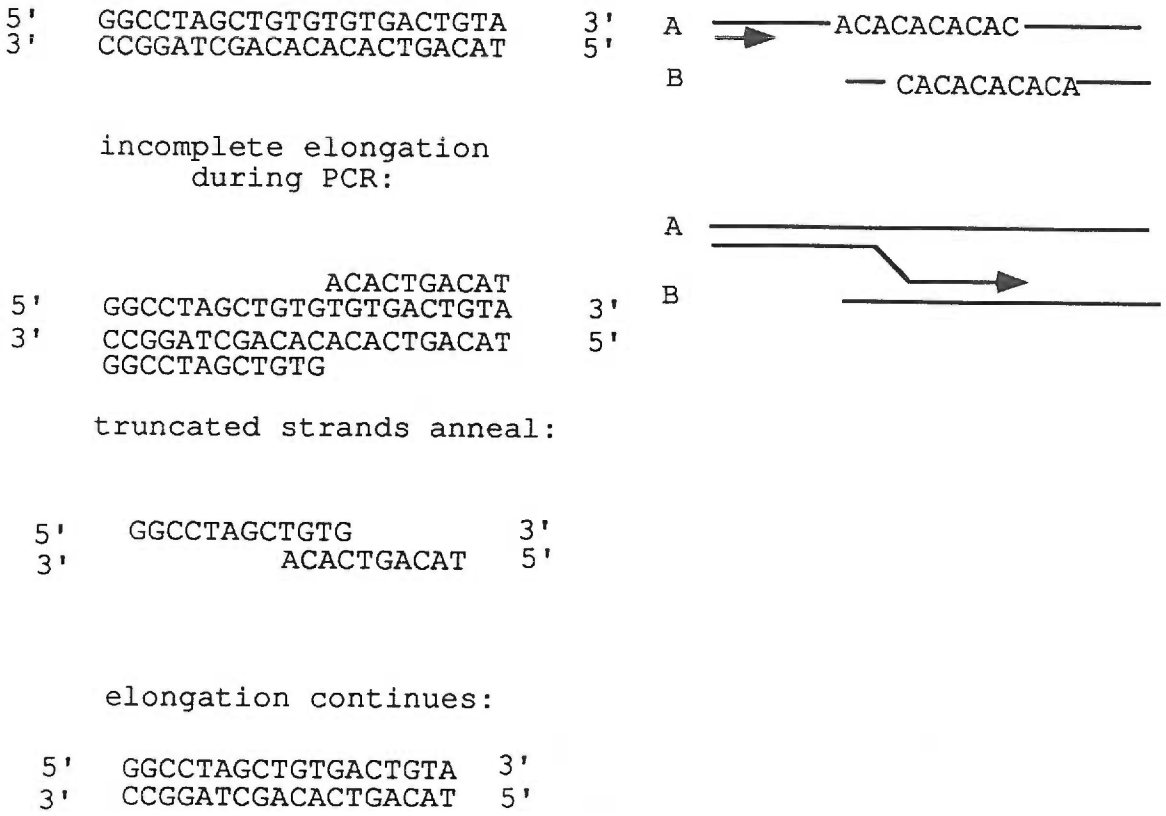


FIG.III-2. Generation of deletions or insertions by annealing of the truncated PCR products between simple sequence tandem repeats and template-switching during the DNA synthesis. Two truncated DNA strands anneal, then elongation continues, giving rise to a 4-base pair deletion in the (CA/GT)n region (left panel). The truncated strand also can bind to the different template based on partial homologous sequence, resulting in deletion or insertion in the (CA/GT)n repeat region (right panel).

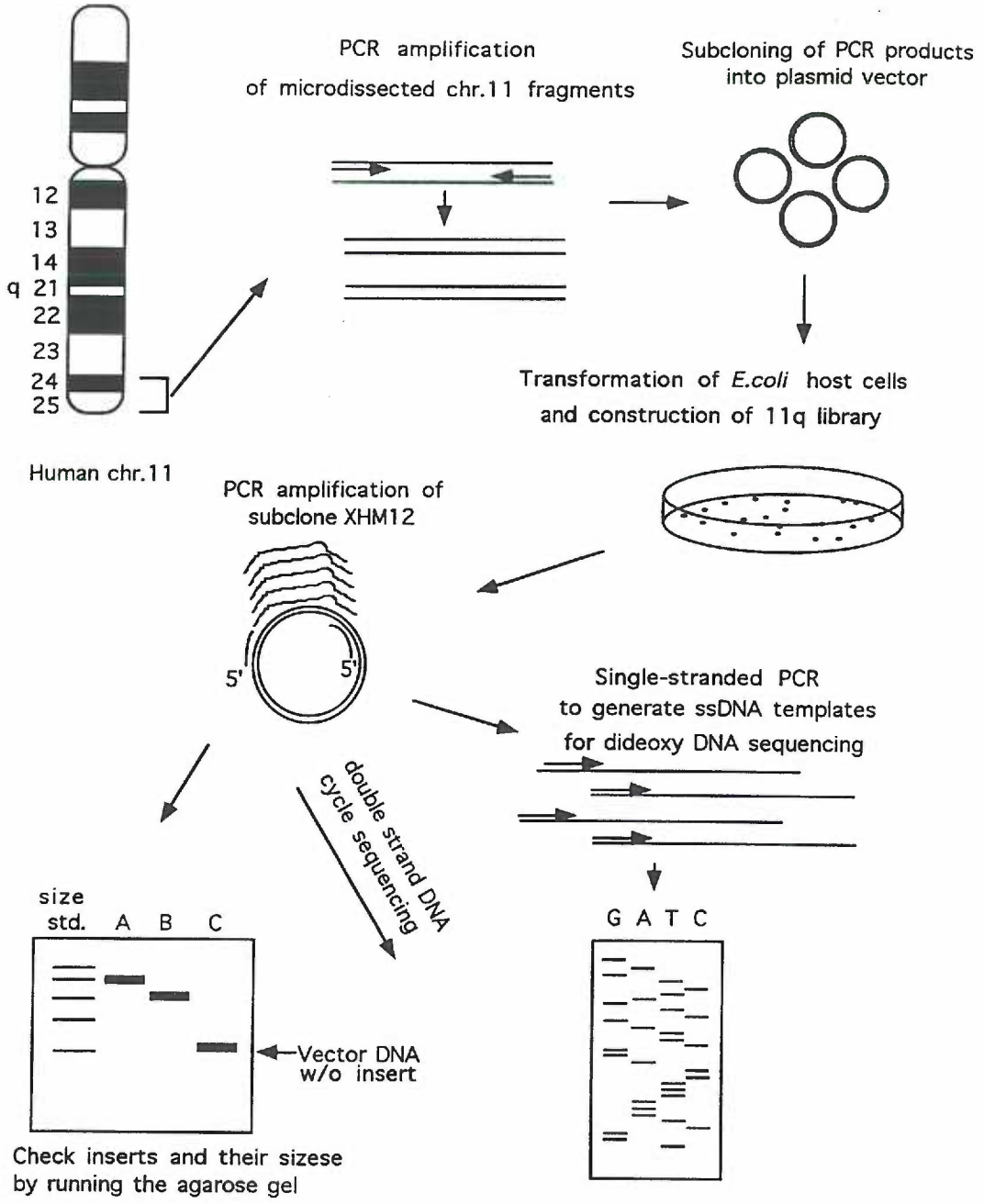
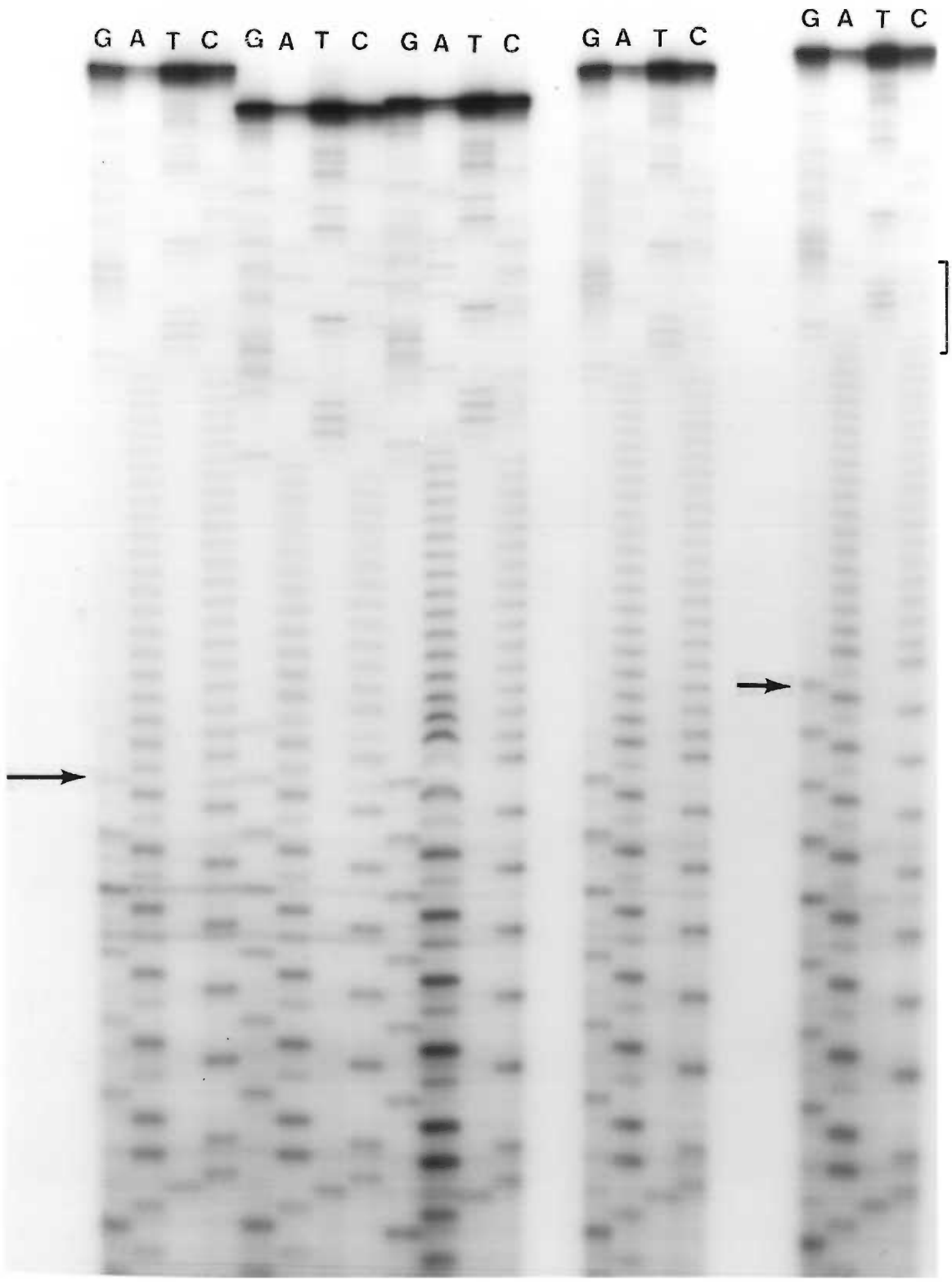


Fig. III-3. Illustration of constructing the 11q library from microdissected 11q24-25 fragments; PCR amplification of insert from a single bacterial colony for checking the size and generating dsDNA for double strand DNA cycle sequencing; single-stranded PCR to generate ssDNA templates for dideoxy DNA sequencing.

Fig.III-4. Molecular basis of polymorphisms in the tetra- and dinucleotide repeat regions at D11S527 detected by using primer jg4A. Sequencing ladders shown here represent 5 unrelated individuals. All five individuals showed various number of repeat units in the dinucleotide repeat region. Three of them have 6 repeat units in the tetranucleotide region (left arrow), which represents 75% randoms examined. Two individuals showed various number of repeat units in the tetranucleotide repeat region (right arrow).


G A T C G A T C G A T C
G A T C
G A T C

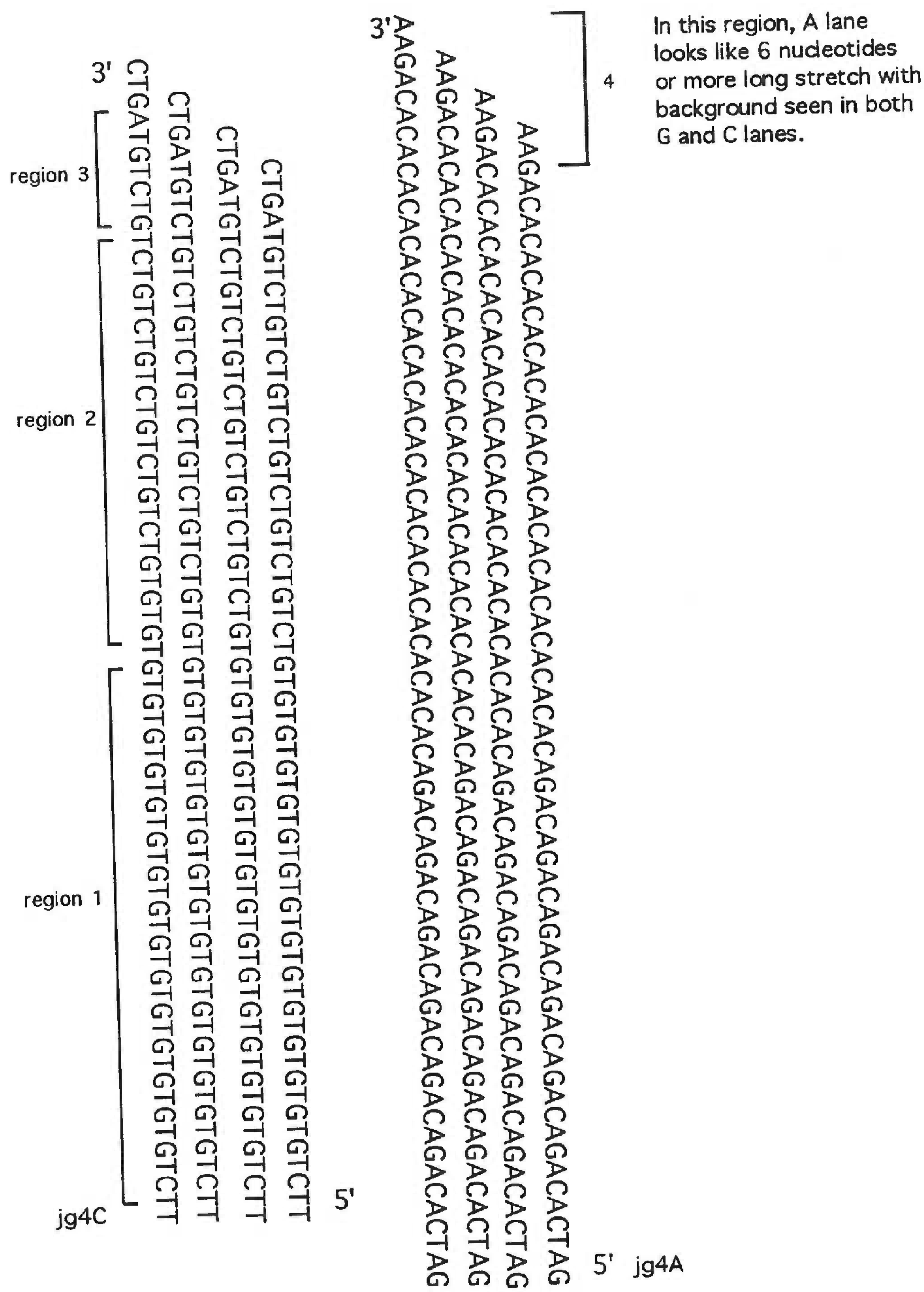
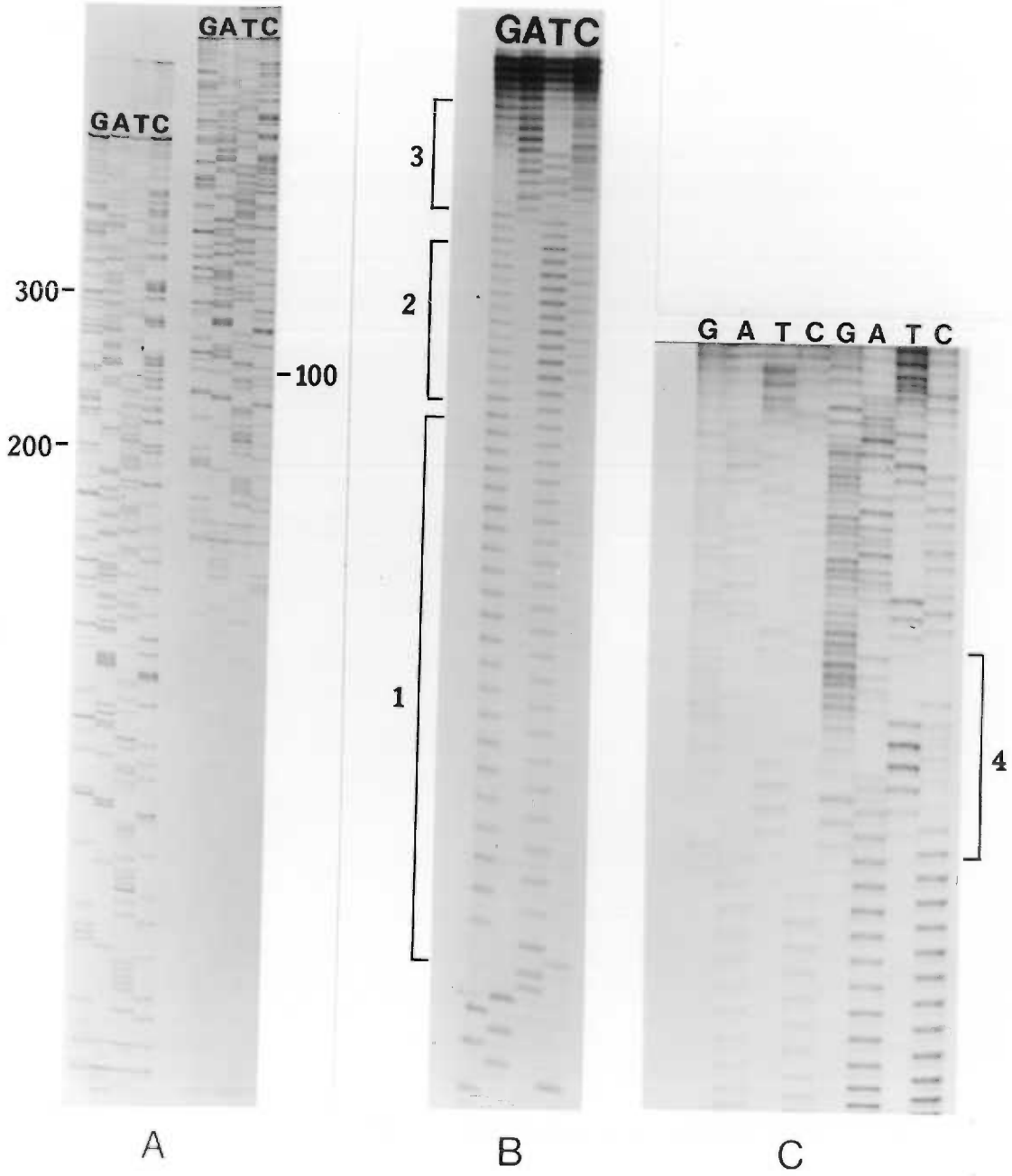


Fig.III-5.
Illustration of deletions or insertions by SSM which may occur in (CA)17 repeat region at the D11S527 locus during PCR. Each DNA fragment is (CA)1 shorter than the one on its left but (CA)1 longer than that on the right. The DNA sequences shown on the left can be divided into 3 regions: 1) dinucleotide repeat region. When using jg4C primer, sequencing ladder in this region is unique and sharp. 2) tetranucleotide repeat region. Likewise, bands in G and C lanes show weaker signals compared with that seen in T lane. Since T appears in every other nucleotide positions in TGTC repeat units, one would expect to see G and C appear at the same position on the sequencing gel if DNA strands are of 2 nucleotide difference in length. 3) the 3' end of tetranucleotide region. The sequence GATC appears as many as 3 or 4 times and bands are seen in all 4 lanes. When using jg4A as the primer, the oppsite was observed. In (TGTC)n repeat, bands are unique and sharp. Toward the 3' end of (CA)n repeat, bands start to cross all 4 lanes.

Fig.III-6. Deletions or insertions which occurred in the (CA)n repeat region at the D11S527 locus during PCR. A. Partial DNA sequence from subclone XHM12 serves as control, for it does not contain any (CA)n repeat. B. Using jg4C primer, the sequencing ladder is unique in the (CA)n repeat region but ambiguous in the (TGTC)n repeat region in which G and C lanes appear at the same positions. C. Using jg4A primer, the sequencing ladder is unique in both tetra- and 5′ portion of dinucleotide repeat regions but ambiguous at the end of dinucleotide repeat, which is just the opposite of what we have seen in B. The entire sequence of this compound repeat is shown in Fig III-4. The 3′ portion of dinucleotide repeat is marked (]) in C and Fig III-4.


GATC
GATC
300-
200-
-100
A
GATC
3
2
1
B
G A T C G A T C
4
C

```
nt
  1 AAAAATTTGG TAACTGCTTT GTAAATGCGG CTCCAAGACA AGTTCTTATC

 51 CAGCCTGTCC TGTCCTGAGC CCCTCCTGCA GAGATTCTGT GTGTGTGTGT
                                          *** **********
101 GTGTGTGTGT GTGTGTGTGT GTCTGTCTGT CTGTCTGTCT GTCTGTGATC
    ********** ********** ********** ********** ******
151 TCATCTCCAG ACAAGTAGAG GGGCAC
```

Fig.III-7. The target sequence flanked by primers jg4C and jg4A. Two PCR primers are underlined. Primer jg4C starts at 25th nucleotide and primer jg4A starts at 155th nucleotide. The target sequence is marked by asterisks.

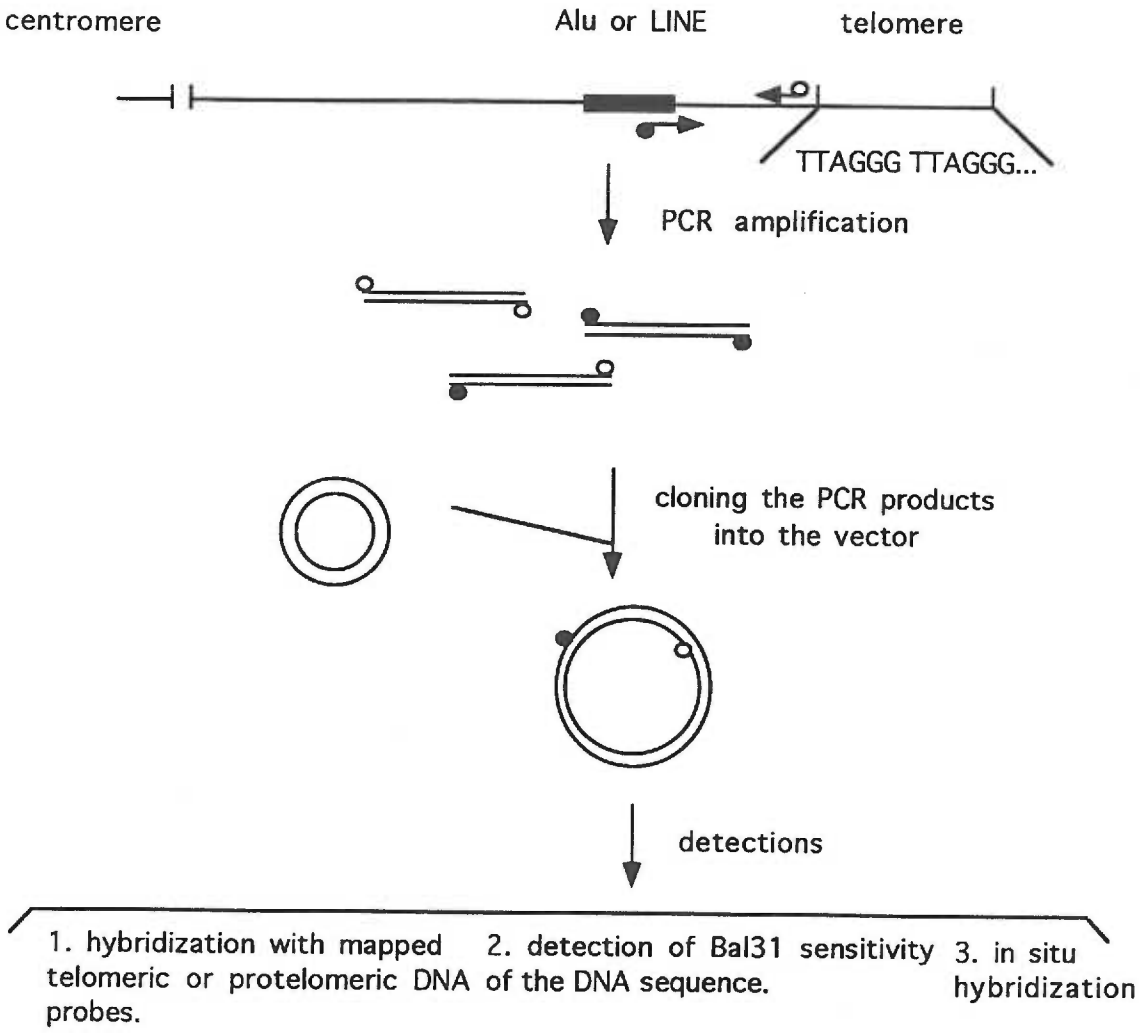


Fig.IV-1. Isolation of polymorphic markers from the telomere region of human chromosome 11q by protelomere-INS PCR. Different restriction enzyme cleavage sites at the 5' ends of PCR primers are shown as small circles attached to arrows. Since the vector is digested with 2 different enzymes simultaneously, only those fragments with different ends will be cloned into it.

Fig.IV-2.

A. PCR amplification of DNAs from 4 somatic cell hybrids, a mouse cell line and human using a single Alu primer. In Lanes 1 to 6 on the left, Alu primer v559pst was used. In lanes 1 to 6 on the right Alu primer v517 was used. Lane 1: J1-44; lane 2: MC-1; lane 3: TG5D1-1; lane 4:J1-11; lane 5: RAG; lane 6: human genomic DNA. The 123 bp DNA ladder served as the size standard (BRL).

B. PCR products amplified by using primers v559pst and vpTH2 were purified , digested with restriction enzymes and subcloned into the plasmid vector pTZ18u. Eight white colonies were randomly picked up from agar plates and grown overnight. The inserts were amplified by using M13 universal and reverse primers. Note that the size of inserts ranges from 400 bp to 2 kb. The 123 bp DNA ladder served as the size standard.

C. PCR amplification of DNAs from somatic cell hybrid J1-44 (lanes 3 and 6), RAG (lanes 1 and 4) and E36 (lanes 2 and 5) by using primers v559pst and vpTH2. Note that the PCR products with sizes of about 150 bp to 1-2 Kb had been generated from J1-44, but not from 2 rodent cell lines. The 123 bp DNA ladder served as the size standard.

D. PCR amplification of DNAs from 4 somatic cell hybrids and rodent cell lines by either using v559pst and vpTH2 (left) or v559pst (right) alone. PCR was done at the higher annealing temperature. Note that the background seen in the rodent cell lines in (A) has been reduced substantially. The 123 bp DNA ladder served as the size standard.

Left panel: v559pst,vpTH2
1).J1-11
2).MC-1
3).J1-44
4).TG5D1-1
5).RAG(mouse)
6).E36(hamster)

Right panel:v599pst
1).MC-1
2).J1-11
3).J1-44
4).TG5D1-1
5).RAG(mouse)
6).E36(hamster)

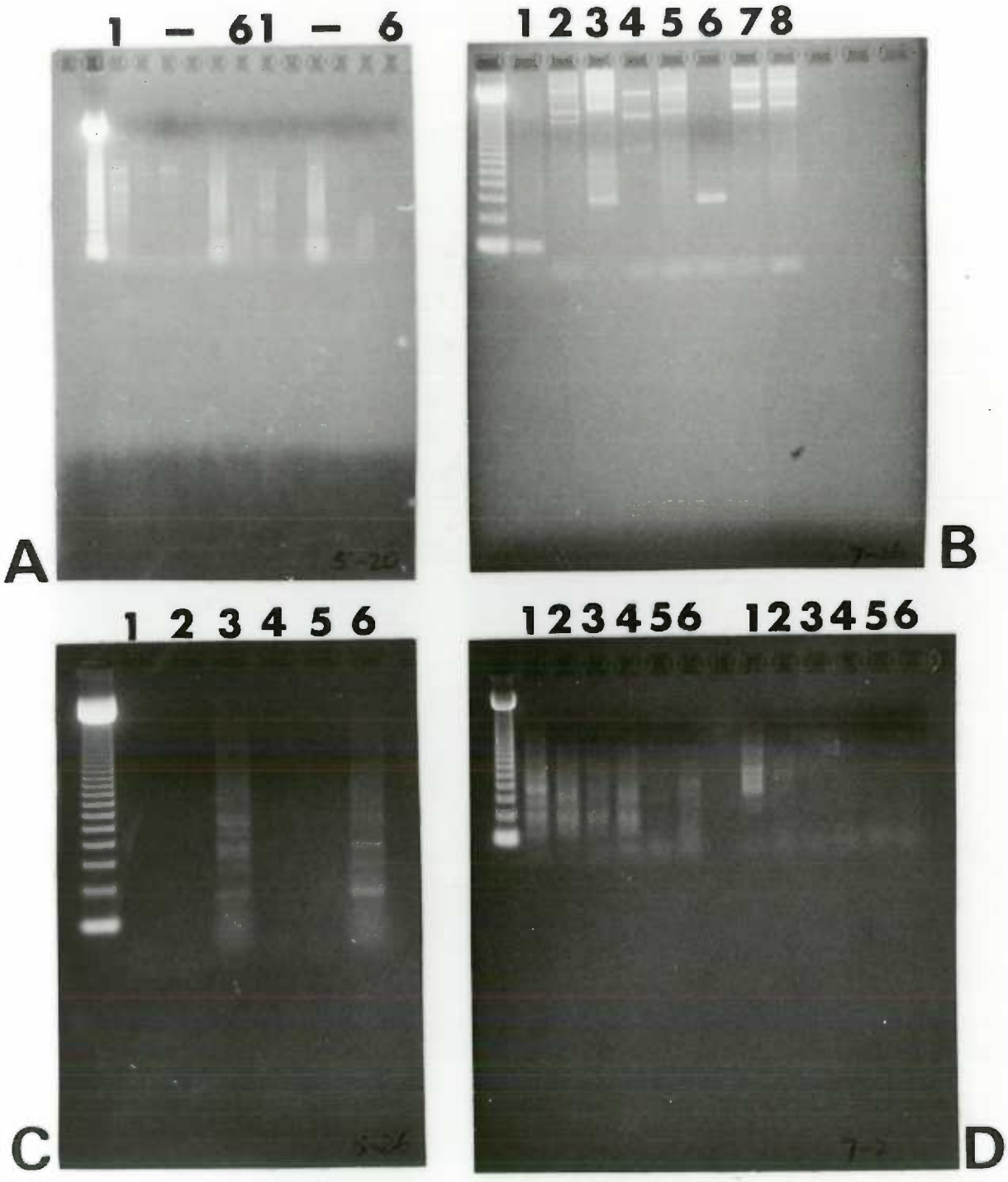

1 – 61 – 6
A
1 2 3 4 5 6 7 8
B
1 2 3 4 5 6
C
1 2 3 4 5 6 1 2 3 4 5 6
D

Table I-1.
The Summary of Fifty Human Chromosome 11q Cosmids

Cosmid	(CA)n Positive	No.of Repeat Block(s)	Sequenced	Chromosomal Localization
13/1	+/(AC)17	1	Yes	q22.3-q23
8/5	+/(GT)18	1	Yes	q23.3-q24
1/16	+/(GT)14	2	Yes	q23.3
9/4	+/()	1	Yes	q23.3
22/8	+-/(AC)9	1	Yes	q
30/1	+/(AC)19	1	Yes	q13
20/13	+-/(AC)10	1	Yes	q
23/2	-			q25
24/10	-			q
3/27	-			q
16/7	-			q
4/3	-			q
12/16	-			q
17/15	-			q
31/4	-			q
5/8	-			q23.3
8/17	-			q
20/7	-			q
25/3	-			q
4/7	-			q
24/25	+/(CA)16	1	Yes	q22.3
1A1(1.1)	-			q13.3-q13.5
9B5(ZD8)	-			q13.3-q13.5
2G9(7.21)+/ 8G11		2		q13.5
(23.23)	+/(GT)17(CTGT)6		Yes	q13.5
9A2 (17.26)	+/(GT)17(CTGT)6		Yes	q13.5
6A3 (9.27)	+/(CA)23,18	5	Yes	q14.1
2C4 (3.16)	-			q21
C5,35	+/(AC)13	1	Yes	q13
C2,11	+/(AC)13N(CA)23		Yes	q24-q25
C7,24	-			q24-q25
C20,22	-			q24-q25
C3,17	-			q24-q25
11-36[a)]	+/	1	Yes	q24-q25
11-270	-			q13.5-q14
11-46	-			q24-q25
11-245	-			q24-q25
11-346	-			q24-q25
11-365	-			q24-q25
11-371	-			q24-q25
11-455	-			q24-q25
11-498	-			q24-q25
11-503	+-	1	No	q24-q25

Table I-1. cont.

11-578	-			q24-q25
11-325	-			q21-q22
11-415	+/(CA)24	1	Yes	q24-25
11-539	+-/(AC)6N24 (CA)3N18(CA)3	1	Yes	q24-25
Phi2-25	+-/(GT)8	1	Yes	q22
L962[b)]	+-/	1	Yes	q24-q25
L605	+-		Yes	q24-q25

a).The full name should be cCI11-.
b).The full name should be CRI-L962 and CRI-L605.

Table I-2. The summary of (GT/CA) dinucleotide repeats in fourteen 11q cosmids

Locus	Cosmid/Phage	Heter.	Localization	No. of repeat	PCR primer sequence
DRD2	24/25	0.76 [1]	q22-q23	(CA)16	CAGGAGCACGTTTCTCTAC
					GGAGGGCGGTGCGGTCAT
D11S420	8/5	0.66	q22.3	(GT)18	AGTTACACCGGTTCTGCAGA
					GATTAATGATAGTGCTATCC
D11S528	13/1	0.60	q22.3-q23	(AC)17	AATGGTGTCCCCACACATGT
					TCCTACCTACCGAGCTTAAA
D11S490	1/16	0.72	q23.3	(GT)14	CACAAACATTGGCGCAT
					TTCTGGGTCACGGTGTTCA
D11S534	30/1	0.74	q13	(AC)19	ATATGGAAACTCTCCGTACT
					GCAACCATGGAGAGTCTGGA
D11S439	cCI11-36	0.78	q24-q25	(AC)19	GATAGTGCTATCCTCTTGCA
					GTTACACCGGTTCTGCAGAT
D11S527	8G11(jg4)	0.90	q14	(GT)17(CTGT)6	ATGCGGCTCCAAGACAAGTTC [3]
					GCCCCTCTACTTGTCTGGAG
D11S83	Phi 2-25	0.31	q24-q25	(GT)8	AGCAGTGGGAGTGACCTCT
					CCTCTGGGAGTATAGGAT
	c5,35	0.62	q	(AC)13	GGCTGGAGATGCTACTAATA
					TCTGATAAATGTCACCAGA
	22/8	0	q	(AC)12	CGACAGACACTTGAGGTTTC
					CAAGAGCTGTGCAGGGCCAT
	6A3	0	q14	(CA)23;(CA)18 [2]	TGACATTGACATCTTCCTGC
					TCCCAAGGCTCTTTCATAGAT
	415	0	q	(CA)24	TACGACTGCTATAGGGAT
					AGCACCAATACCAGAGGTCT
	2,11	0	q	(AC)13/(CA)23	CCAGCCCAATAATCATTGCA
					CTTTCTTGCCTCTCCATCGA
	9A2	0	q13	(CA)17	ACAAGTTCTTATCCAGCCTG
					GGGATCCTCTAGAGTCGATC

1). Combined PIC, otherwise is 0.62.
2). There are two seperate (CA)n repeat found in cosmid 6A3.
3). PCR primers for D11S527 were designed by David Browne et al. (1991).

Table I-3.
The Summary of Human Chromosome 11q (CA)n Repeat Loci

Locus (Probe)	Alleles(bp)	Freq.	Het.	Ref.
DRD2 (24/25)	86	0.15	0.76(138)[1]	Hauge et al
	84	0.47		
	82	0.22		
	80	0.16		
D11S420 (8/5)	208	0.063	0.66(66)[2]	Luo et al
	196	0.063		
	194	0.063		
	192	0.063		
	190	0.36		
	188	0.39		
D11S528 (13/1)	91	0.02	0.60(56)	Hauge et al
	89	0.02		
	87	0.02		
	83	0.05		
	81	0.02		
	79	0.04		
	77	0.23		
	75	0.05		
	73	0.55		
D11S490 (1/16)	167	0.01	0.72(100)	Luo et al
	165	0.07		
	163	0.12		
	161	0.07		
	159	0.09		
	157	0.43		
	155	0.01		
	153	0.02		
	147	0.18		
D11S534 (30/1)	244	0.008	0.74(116)	Hauge et al
	242	0.05		
	240	0.15		
	238	0.34		
	236	0.10		
	234	0.28		
	232	0.03		
	230	0.03		
	228	0.012		
D11S527(8G11/jg4)	166	0.013	0.90(154)	Browne et al
	164	0.11		
	162	0.097		
	160	0.156		
	158	0.162		
	156	0.058		
	154	0.136		
	152	0.019		
	150	0.091		

Table I-3. cont.

	148	0.078		
	146	0.071		
	142	0.006		
D11S439 (cCI11-36)	188	0.016	0.78(160)	Hauge et al
	186	0.016		in press
	182	0.040		
	180	0.030		
	176	0.016		
	170	0.060		
	168	0.051		
	166	0.030		
	164	0.37		
	162	0.34		
	160	0.016		
5,35	A1	0.035	0.62(58)	
	A2	0.104		
	A3	0.017		
	A4	0.19		
	A5	0.017		
	A6	0.57		
	A7	0.052		
	A8	0.017		
D11S83 (Phi2-25)	A1	0.22	0.31(32)	
	A2	0.78		

1). Combined PIC of (CA)n repeat and Taq RFLPs at this locus, otherwise the heterozygosity of (CA)n repeat it is 0.62.

2). The number of chromosome analized is given in parentheses.

Table I-4. Chi Square Test.

Chr.Location	% of cosmids tested	No.of Posi./ No.of Tl. Posi.	O	E	$(O-E)^2/E$
q13,q22-q23 q14,q21,	49(18/37)	8/12	8	6	.67
q24-q25	51(19/37)	4/12	4	6	.67
					$X^2=1.34$
					df=1,alpha=.25

Table I-5. Markers used for construction of the index map of human chromosome 11q

Locus	Probe	% Het.	Type of polymorphism	Reference
D11S534	c30,1	74	(CA)n repeat	Hauge et al.1991
D11S533	4F7	88	compound hexanucleotide repeat	Eubanks et al. 1991
sMSH3	sMSH3/pcr	82	(CA)n repeat	Donis-Keller pers.comm.
PYGM	sHI-1/pcr	89	(CA)n(GA)n repeat	Donis-Keller pers.comm.
INT2	INT2/pcr	77	(CCAT)n repeat	Donis-Keller pers.comm.
D11S288/ D11S436	cCI11-30/ p3C7	74	haplotyped RFLPs	Leach et al.1984*
D11S388	CJ52.4	88	3 Msp RFLPs haplotyped	Julier et al. 1990*
DRD2	c24/25/ hD2G1	76	(CA)n repeat and Taq RFLPs haplotyped	Hauge et al.1991; Grandy et al. 1989b
D11S144/ D11S351	MCT128.1/ CJ52.208	79	haplotyped RFLPs	Carlson et al.1988b*
D11S490	c1,16	72	(CA)n repeat	Luo et al. 1990
D11S527	jg4 (8G11)	90	(GT)17(CTGT)6 repeat	Brown et al. 1991
D11S35	phage-2-22	86	(CA)n repeat and Taq RFLP haplotyped	Litt et al .(HGM10); McClure et al .1986
D11S97	pms51	77	VNTR	Royle et al. 1987*
D11S439	cCI11-36	78	(CA)n repeat	Hauge et al . in press
Mfd231	Mfd231/pcr	84	(CA)n repeat	Weber pers.comm.
D11S614	6d11/pcr	80	(CA)n repeat	Weber pers.comm.
sMSH2	sMSH2			Donis-Keller pers.comm.

* Data were obtained from CEPH database V.5.

Table I-6. Two-point linkage analysis of 17 index markers from the long arm of human chromosome 11. (bottom diagonal is recombination; top is LOD score)

o D11S534 1 D11S533 2 D11S614 3 S288/S436 4 D11S388 5 DRD2
6 Mfd231 7 S144/S351 8 D11S490 9 D11S439 10 D11S527 11 D11S35
12 D11S97 13 INT2 14 PYGM 15 sMSH3 16 CD2 17 sMSH2

	0	1	2	3	4	5	6	7	8	9
0		101.98	5.34	30.46	18.39	6.89	14.99	6.76	5.81	1.51
1	0.02		7.61	32.02	28.98	12.47	20.86	8.53	9.68	3.06
2	0.34	0.32		4.90	9.33	29.74	47.96	39.84	66.49	23.64
3	0.08	0.11	0.29		5.84	3.84	6.13	2.71	2.88	1.87
4	0.19	0.16	0.26	0.23		17.31	23.21	9.32	11.11	1.73
5	0.31	0.27	0.10	0.27	0.19		72.96	32.98	39.88	10.66
6	0.26	0.26	0.10	0.29	0.18	0.01		41.52	60.68	19.03
7	0.29	0.28	0.04	0.28	0.22	0.05	0.07		39.03	13.61
8	0.33	0.31	0.02	0.33	0.25	0.08	0.09	0.04		24.12
9	0.41	0.38	0.16	0.34	0.38	0.25	0.22	0.17	0.16	
10	0.02	0.00	0.33	0.12	0.16	0.27	0.25	0.28	0.32	0.42
11	0.23	0.20	0.18	0.26	0.10	0.10	0.09	0.17	0.18	0.30
12	0.06	0.07	0.34	0.05	0.22	0.30	0.30	0.32	0.35	0.40
13	0.03	0.06	0.32	0.05	0.21	0.29	0.28	0.32	0.34	0.42
14	0.09	0.11	0.31	0.03	0.25	0.29	0.29	0.27	0.31	0.42
15	0.10	0.12	0.34	0.04	0.24	0.34	0.30	0.31	0.34	0.41
16	0.31	0.30	0.02	0.29	0.22	0.09	0.10	0.05	0.02	0.14
17	0.20	0.22	0.15	0.18	0.06	0.09	0.09	0.11	0.16	0.39

	10	11	12	13	14	15	16	17
0	94.63	21.60	67.19	72.98	53.04	53.36	5.45	4.59
1	129.05	36.97	74.37	70.77	52.41	57.08	8.06	4.49
2	6.03	27.04	5.36	4.35	6.34	4.82	60.67	9.10
3	26.95	7.55	39.00	38.66	38.05	49.33	2.93	2.84
4	25.23	43.90	16.97	11.97	9.48	13.24	10.24	8.27
5	11.81	43.81	7.69	5.55	6.64	4.03	32.15	6.71
6	19.21	71.07	11.72	9.98	11.50	11.67	51.39	11.32
7	7.60	22.99	7.23	3.31	7.27	5.39	36.35	2.92
8	7.69	32.85	4.99	4.26	7.77	5.77	76.70	6.95
9	1.28	8.31	2.00	0.80	0.99	1.41	28.97	0.63
10		30.73	64.53	58.60	49.40	51.66	7.33	3.64
11	0.20		19.15	11.85	14.46	15.76	21.97	15.04
12	0.07	0.25		76.68	60.32	65.53	2.64	5.02
13	0.06	0.27	0.01		58.60	66.96	4.74	3.49
14	0.10	0.26	0.06	0.05		83.79	6.32	3.17
15	0.11	0.28	0.08	0.06	0.03		5.87	4.37
16	0.30	0.20	0.36	0.30	0.30	0.31		5.92
17	0.24	0.03	0.19	0.23	0.27	0.22	0.13	

References

Bates,G.P., MacDonald,M.E., Baxendale,S., Youngman,S.,Lin,C., Whaley,W.L., Wasmuth,J.J., Gusella,J.F. and Lehrach,H. (1991) Defined Physical Limits of the Huntington Disease Gene Candidate Region. Am.J.Hum. Genet.49:7-16.

Beats,B.J.B.,Sherman,S.L.,Morton,N.E.,Roboson,E.B.,Buetow,K .H.,Cartwright,P.E.,Chakravarti,A.,Francke,U.,Green,P.P.,an d Ott,J. (1991) Guidelines for Human Linkage Maps:An International System for Human Linkage Maps.(ISLM 1990). Genomics:557-560.

Botstein, D., White,R.L., Skolnick,M. and Davis, R.W. (1980) Construction of A Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. Am.J.Hum.Genet. 32: 314-331.

Brown,D.L., Gault, J., Thompson,M., Hauge, X.Y., Evans, G.A. and Litt,M. (1991) Dinucleotide Repeat Polymorphism at the
D11S527 Locus. Nucl. Acids Res. 19:4790.

Budarf,M.,Sellinger,B.,Griffin,C.,and Emanuel,B.S.(1989) Comparative Mapping of the Constitutional and Tumor -associated 11;22 Translocations.
Am.J.Hum.Genet.45:128-139.

Buetow,K. (1991) Influence of Aberrant Observations on High-Resolution Linkage Analysis Outcomes. Am.J.Hum.Gnet. 49: 985-994.

Carlson. M., Nakamura, Y., Gillilan, S., O'Connell, P., Lebo R., Gobin, F., Lathrop, M., Lalouel, J.M and White, R. (1988b) Isolation and Mapping of A Polymorphic DNA Sequence MCT 128.1 on Chromosome 11 [D11S285]. Nucl. Acids Res. 16: 378.

Charron,M.L., Read,E., Haut,S.R. and Michels,C.A. (1988) Molecular Evolution of the Telomere Associated MAL Loci of Saccharomyces. Genetics.122,307-316.

Cheng, J., Smith, C.L. and Cantor, C.R. (1989) Isolation and Characterization of A Human Telomere. Nucl. Acids. Res. 17: 6109-6127.

Clemens,P.R., Frenwick,R.G., Chamberlain,J.S., Gibbs,R.A., de Andrade, M.,Chakraborty,R. and Caskey,T.C. (1991) Carrier Detection and Prenatal Diagnosis in Duchenne and

Becker Muscular Dystrophy Families Using Dinucleotide Repeat Polymorphisms. Am.J.Hum.Genet. 49: 951-60.

Cookson,W., Sharp,P., Faux,J. and Hopkin,J. (1989) Linkage between Immunoglobulin E Responses Underlying Asthma and Rhinitis and Chromosome 11q. Lancet. 1: 1292-1295.

Cotter,F.E., Hampton,G.M., Nasipuri,S., Bodmer,W.F., and Young,B.D. (1990) Rapid Isolation of Human Chromosome-Specific DNA Probes from A Somatic Cell Hybrid. Genomics. 7: 257-263.

de Lange, T., Shine, L., Myers, R.M., Cox, D.R., Naylor, S.L., Killery, A.M., and Varmus, H.E. (1990) Structure and Variability of Human Chromosome Ends. Molecular and Cellular Biology. 10: 518-527.

de Taisne,C., Gegonne,A., Stehelin,D., Bernheim,A. and Berger,R. (1984) Chromosomal Localization of the Human Proto-Oncogene c-est. Nature. 310: 581-583.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K and Mattick, S.J. (1991) "Touchdown" PCR To Circumvent Spurious Priming During Gene Amplification. Nucl.Acids. Res. 19:4008.

Donis-Keller,H., Green,P., Helms,C., Cartinhour,S., Weiffenbach,B., Stephens,K., Keith,T.P., Bowden,D.W., Smith,D.R., Lander,E.S., Bostein,D., Akots,G., Rediker,K.S., Grevius,T., Brown,V.A., Rising,M.B., Parker,C., Powers,J.A., Watt,D.E.,Kanffman,E.R. ,Bricker,A. ,Phipps,P.,Muller-Kahle, H.,Fulton ,T.R.,Ng,S.,Schumm, J.W.,Braman, J.C.,Knowlton, R.G.,Barker, D.F.,Crooks, S.M.,Lincoln,S.E.,Daly,M.J.,and Abrahamson,J.(1987) A Genetic Linkage Map of the Human Genome. Cell. 51:319-337.

Edwards,A., Civitello,A., Hammond,H.A. and Caskey, T. (1991) DNA Typing and Genetic Mapping with Trimeric and Tetrameric Tandem Repeats. Am.J.Hum.Genet. 49: 746-756.

Eubanks, J.H., Selleri, L., Hart, R., Rosette, C. and Evans, G.A. (1991) Isolation, Localization, and Physical Mapping of A Highly Polymorphic Locus on Human Chromosome 11q13. Genomics. 11: 720-729.

Evans,G.A. and Lewis,K.A. (1989) Physical Mapping of Complex Genomes by Cosmid Multiplex Analysis. Proc.Natl.Acad.Sci. 86: 5030-5034.

Feinberg,A.P., and Vogelstein,B. (1983) A Technique for Radiolabeling DNA Restriction Endonuclease Fragments to High Specific Activity. Anal.Biochem. 132: 6-13.

Freund,A., Bichara,M. and Fuchs,R.P.P. (1989) Z-DNA-Forming Sequences Are Spontaneous Deletion Hot Spots. Proc. Natl. Acad. Sci. 86: 7465-7469.

Friend,S.H., Bernards,R., Rogelj, S., Weinberg,R.A., Rapapot,J.M., Albert,D.M. and Dryja,T.P. (1986) A Human DNA Segment with Properties of the Gene That Predisposes to Retinoblastoma and Osteosarcoma. Nature. 323: 643-646.

Fujita,R.,Hanauer,A.,Sirugo,G.,Heiling,R. and Mandel,J.L. (1990) Additional Polymorphisms at Marker Loci D9S35 and D9S15 Generate Extended Haplotypes in Linkage Disequilibrium with Friedreich Ataxia. Proc. Natl. Acad. Sci. USA. 87: 1796-1800.

Gatti,R.A., Berkel,I.,Boder,E., Braedt,G., Charmley,P., Concannon,P., Ersoy, F., Foroud,T., Jaspers,N.G.J., Lange,K., Lathrop,G.M., Leppert,M., Nakamura,Y., O'Connell,P., Patterson,M., Salser,W., Sanal,O.,Silver,J., Sparkes,R.S., Susi,E., Weeks, D.E., Wei,S., White,R. and Yoder, F. (1988) Localization of An Ataxia-Telangiectasia Gene to Chromosome 11q22-23. Nature. 336: 577-580.

Goate, A., Chartier-Harlin,M.-C. Mullan,M., Brown, J., Crawford,F., Fidani,L., Giuffra,L., Haynes,A., Irving, N., James,L., Mant, R., Newton, P., Rooke, K., Roques,P., talbot,C., Pericak-Vance,M., Roses,A., Williamson,R., Rossor,M., Owen,M. and Hardy.J. (1991) Segregation of A Missense Mutation in the Amyloid Precursor Protein Gene with Familial Alzhemer's Disease. Nature. 349: 704-706.
Grandy,D.K., Litt,M.,Allen,L. Bunzow,J.R., Marchionni,M., Makam,H., Reed,L., Magenis,R.E.and Civelli,O. (1989) The Human Dopamine D2 Receptor Gene Is Located on Chromosome 11 at q22-q23 and Identifies a TaqI RFLP. Am.J.Hum.Genet. 45: 778-785.

Han,J.,Lu,C., Brown,G. and Rado,T.A.,(1991) Direct Amplificaton of A Single Dissected Chromosomal Segment by Polymerase Chain Reaction: A Human Brain Sodium Channel Gene Is on Chromosome 2q22-q23. Proc.Natl.Acad.Sci. USA. 88:335-339.

Hang,T., Mak,K. and Fong,K. (1990) Specificity Enhancer for Polymerase Chain Reaction. Nucl. Acids. Res. 18: 4953.

Hanahan,D. (1985) Techniques for Transformation of E. coli. In DNA Cloning : A Practical Approach (ed.D.M.Glover) IRS.Press,Oxford. Vol 1: 109-135.

Harley,H.G., Brook,J.D., Rundel,S.A., Crow,S., Reardon,W., Buckler,A.J., Harper,P.S., Housman,D.E. and Shaw, D.J. (1992) Expansion of An Unstable DNA Region and Phenotypic Variation in Myotonic Dystrophy. Nature. 355: 545-546.

Hauge, X.Y., Evans, G.A. and Litt, M. (1991) Dinucleotide Repeat Polymorphism at the D11S534 Locus. Nucl. Acids Res. 19:4308.

---------- (1991) Dinucleotide Repeat Polymorphism at the D11S528 Locus. Nucl. Acids Res. 19:1964.

Huage XY, Grandy DK, Eubanks JH, Evans GA, Civelli, O. and Litt M. (1991) Detection and Characterization of Additional DNA Polymorphisms in the Dopamine D2 Receptor Gene. Genomics. 10: 527-530.

Hazan,J., Dubay,C., Pankowiak, M., Becuwe, N.and Weissenbach,J. (1992) A Genetic Linkage Map of Human Chromosome 20 Composed Entirely of Microsatellite Markers. Genomics. 12: 183-189.

Hermanson,G.G. Lichter,P., Selleri, L. Ward,D.C. and Evans,G.A. (1992) Cosmid Linking Clones Localized to the Long Arm of Human Chromosome 11. Genomics. 13:134-143.

Hori,T.-A., Takahashi.E.I., Tanigami,A., Tokino,T. and Nakamura, Y. (1992). A High-Resolution Cytogenetic Map of 168 Cosmid DNA Markers for Human Chromosome 11. Genomic. 13: 129-133.

Huang,T.H.-M., Hejtmancik,F., Edwards,Al., Pettigrew, A.L. Herrera, C.A., Hammond, H.A., Caskey, C.T. Zoghbi, H.Y. and Ledbetter, D.H. (1991) Linkage of the Gene for An X-linked Mental Retardation Disorder to A Hypervariable (AGAT)n Repeat Motif within the Human Hypoxanthine Phosphoribosyltransferase (HPRT) Locus (Xq26). Am. J. Hum. Genet. 49: 1312-1319.

Hudson,T., Engelstein,M., Lee,M.K., Ho,E.C., Rubenfield,M.J., Adams,C.P., Housman,D.E. and Dracopoli,N.C. (1992) Isolation and Chromosomal Assignment of 100 Highly Informative Human Simple Sequence Repeat Polymorphisms. Genomics. 13:622-629.

Hyer,R.N.,Julier,C.,Buckley,J.D.,Trucco,M.,Rotter,J.,Spielman,R.,Barnett,A.,Bain,S.,Boitard,C.,Deschamps,I.,Todd,J.A.,Bell,J.I. and Lathrop,G.M. (1991) High Resolution Linkage Mapping for Susceptibility Genes in Human Polygenic Disease: Insulin- dependent Diabetes Mellitus and Chromosome 11q. Am.J.Human.Genet. 48: 243-257.

Iizuka,M., Mashiyama, S., Oshimura, M., Sekiya,T.and Hayashi,K. (1992) Cloning and Polymerase Chain Reaction-Single-Strand Confirmation Polymorphism Analysis of Anonymous Alu Repeats on Chromosome 11. Genomics. 12:139-146.

Ijdo,J.W. Baldini,A., Ward,D.C., Reeders,S.T. and Wells,R.A. (1991) Origin of Human Chromosome 2: An Ancestral Telomere-telomere Fusion. Proc. Natl.Acad. Sci. UAS. 88: 9051-9055.

Iwasaki, H., Stewart,P.W., Dilley, W.G., Holt, M.S., Steinbrueck, T.D., Wells, S. A., Jr., and Donis-Keller, H. (1992) A Minisatellite and A Microsatellite Polymorphism within 1.5 Kb at the Human Muscle Glycogen Phosphorylase (PYGM) Locus Can Be Amplified by PCR and Have Combined Informativeness of PIC 0.95. Genomics. 13: 7-15.

Julier,C., Nakamura,Y., Lathrop.M., O'Connell,P., Leppert,M.,Litt,M. and Mohandas,T.(1990) Detailed Map of the Long Arm of Chromosome 11. Genomics.7: 335-345.

Jeffreys,A.J.,Royle,N.J., Wilson,V. and Wong, Z.(1988) Spontaneous Mutation Rates to New Length Alleles at Tandem-Repetitive Hypervariable Loci in Human DNA. Nature. 332: 278-281.

Jeffreys,A.J.,Wilson,V. and Thein,S.L.(1985) Indivadual-Specific "Fingerprint" of Human DNA. Nature. 316: 76-79.

Jeffreys,A.J., Wilson,V. and Thein,S.L. (1985) Hypervariable "Minisatellite" Regions in Human DNA. Nature. 314: 67-73.

Jeffreys,A.J., Brookfield,J.F.Y. and Semeonoff,R. (1985) Positive Identification of An Immigration Test-Case Using Human DNA Fingerprints.Nature. 317: 818-819.

Jones,C., Kimmel,K.A., Carey,T.E., Miller,Y.E., Lehman,D.W. and MacKenzie,D. (1983) Furhter Studies on A Hybrid Cell-Surface Antigen Associated with Human Chromosome 11 Using a Monoclonal Antibody. Somatic Cell Genetics. 9:489-496.

Jones,C., Bill, J., Larizza,L., Pym,B., Goodfellow,P. and Tunnacliffe,A. (1984) Relationships Between Genes on Human Chromosome 11 Encoding Cell-Surface Antigens. Somatic cell and Molecular Genetics. 10:423-428.

Kaltenboeck, B., Spatafora, J.W.,Zhang, X., Kousoulas, K.G., Blackwell, M. and Storz, J. (1992) Efficient Production of Single-Stranded DNA as Long as 2 kb for Sequencing of PCR-Amplified DNA. Biotechniques: 164-171.

Kao,F.and Yu,J.(1991) Chromosome Microdissection and Cloning in Human Genome and Genetic Disease Analysis. Proc.Natl.Acad.Sci. USA. 88: 1844-1848.

Kariya,Y., Kato,K., Hayashizaki,Y., Himeno,S., Tarui,S. and Matsubara,K. (1987) Revision of Consensus Sequence of Human Alu Repeats--A Review. Gene. 53:1-10.

Kinzler,K.W., Nilbert,M.C., Su,L.-K. Vogelstein,B., Bryan,T.M., Levy,D.B., Smith,K.J., Preisinger,A.C., hedge,P., mckechnie,D., Finniear,R., markham, A., Groffen,J., Boguski,M.S., Altschul, S.F., Horri,A., Ando,H., Miyoshi,Y., Miki,Y., Nishisho, I. and Nakamura,Y. (1991) Identification of FAP Locus Genes from Chromosome 5q21. Science. 253: 661-665.

Kunkel,L.M., Monaco,A.P. middlesworth,W., Ochs,H.D. and latt, S.A. (1985) Specific Cloning of DNA Fragments Absent from the DNA of A Male Patient with An X Chromosome Deletion. Pro. Natl. Acad. Sci. 82: 4778-4782.

Krishnan, B.R., Kersulyte, D., Brikun, I., Berg, C.M. and Berg, D.E. (1991) Direct and Crossover PCR Amplification To Facilitate Tn5supF-Based Sequencing of lambda Phage Clones. Nucl. Acids Res. 19:6177-6182.

Larsson,C., Skogseid,B., Nakamura,Y. and Nordenskjold,M. (1988) Multiple Endocrine Neoplasia Type I Gene Maps To Chromosome 11 and Is Lost in Insulinoma. Nature. 332: 85-87.

Lathrop, G.M., Lalouel.J.M., Julier,C. and Ott,J. (1985) Multilocus Linkage Analysis in Humans: Detection of Linkage and Estimation of Recombination. Am.J.Hum.Genet. 37: 482-498.

-------- (1984) Strategies for Multilocus Linkage Analysis in Humans. Proc.Natl.Acad.Sci. USA. 81:3443-3446.

Leach, R., Cavenee, W. and White, R. (1984) Isolation of Polymorphic DNA Sequences from Chromosome 6. Cytogenet Cell Genet. 37: 521.

Ledbetter,S.A., Nelson,D.L., Warren,S.T. and Ledbetter,D.H. (1990) Rapid Isolation of DNA Probe within Specific Chromosome Regions by Interspersed Repetitive Sequence Polymerase Chain Reaction. Genomics. 6: 475-481.

Leppert,M.,O'Connell,P., Nakamura,Y., Lathrop,M., Maslen,S., Litt, M., Cartwright,P., Lalouel,J.-M. and White,R. (1987) A Partial Primary Genetic Linkage Map of Chromosome 11. Human Gene Mapping 9. Cytogenet.Cell Genet. 46:648.

Levinson,G. and Gutman,G.A. (1987a) High Frequencies of Short Frameshifts in Poly-CA/TG Tandem Repeats Borne by Bacteriophage M13 in Escherichia coli K-12. Nucl.Acids Res. 15: 5323-5338.

Levinson,A. and Gutman,G.A. (1987b) Slipped-Strand Mispairing: A Major Mechanism for DNA Sequence Evolution. Mol. Biol. Evol. 4: 203-221.

Lichter,P.,Tang,C.C.,Call,K.,Hermanson,G.,Evans,G.A., Housman,D.and Ward,D.C. (1990) High-Resolution Mapping of Human Chromosome 11 by in situ Hybridization with Cosmid Clones. Science. 24: 64-69.

Litt, M.,Sharma, V. and Luty, J.A. A highly Polymorphic (TG)n Microsatellite at the D11S35 Locus. (HGM 10) [A2053].

Litt,M., Bruns,G.A., Sheehy,R. and Magenis,R.E. (1986) A Highly Polymorphic Locus in Human DNA Revealed by Probes from Cosmid 1-5 Maps to Chromosome 2q35-37. Am. J. Hum. Gene. 38: 288-296.

Litt,M. and Luty J.A. (1989) A Hypervariable Microsatellite Revealed by In Vitro Amplification of A Dinucleotide Repeat within the Cardiac Muscle Actin Gene. Am.J. Hum.Genet. 44,397-401.

Litt,M. (1991) PCR of Microsatellite. in" PCR, a Practical Approach". IRL Press. Oxford.

Luo, X.Y., Evans, G.A. and Litt, M. (1990) Dinucleotide Repeat Polymorphism at the D11S490 Locus. Nucl. Acids Res. 18: 7470.

---------- (1990) Dinucleotide Repeat Polymorphism at the D11S420 Locus. Nucl. Acids Res. 18: 5920.
Luty,J.A., Guo,Z.,Willard,H.F., Ledbetter,D.H. Ledbetter,S. and Litt,M. (1991) Five Polymorphic Microsatellite VNTRs on the Human X Chromosome. Am. J. Hum. Genet. 46: 776-783.

LutzelSchwab,R., Muller,G., Walder,B.,Schmidt,O.,Furbass,R. and Mechler,B. (1986) Insertion Mutation Inactivates the Expression of the Recessive Oncogene Lethal(2) Giant Larvae of Drosophila melanogaster. Molec. Gen. Genet. 204: 58-63.

Marton, A., Delbecchi, L. and Bourgaux, P. (1991) DNA Nicking Favors PCR Recombination. Nucl.Acids. Res. 19:2423-2426.

Maslen,C.L., Jones,C., Glaser,T., Magenis,R.E., Sheehy,R., Kellogg,J. and Litt,M. (1988) Seven Polymorphic Loci Mapping To Human Chromosomal Region 11q22-qter. Genomics. 2: 66-75.

McClure, C.M., Bufton, L., Mohondas, T.K., Sheehy, R., Magenis,R.E. and Litt, M. (1986) RFLPs Revealed by Probes From Genomic Libraries Specific for Chromosome Regions 11q and 16q. Am. J. Hum. Genet. 39: A 162 [H3596].

Mckusick, V.A. (1990) Mendelian Inheritance in Man. The Johns Hopkins University Press. Baltimore.

---------- (1987) The Morbid Anatomy of the Human Genome : A Review of Gene Mapping in Clinical Medicine. Medicine. 66: 237-296.

Meyerhans, A., Vartanian, J-P. and Wain-Hobson,S. (1990) DNA Recombination During PCR. Nucl. Acids. Res. 18:1687-1691.

Morton, N. (1991) Parameters of the Human Genome. Proc. Natl. Acad. Sci.UAS. 88:7474-7476.

Nakamura,Y., Leppert,M., O'Connell P., Wolff,R., Holm,T., Culver,M., Martin,C., Fujimoto,E., Hoff,M., Kumlin,E. and White,R. (1987) Variable Number of Tandem Repeat (VNTR) Makers for Human Gene Mapping. Science. 235: 1616-1622.

Namba,H., Naranara, K., Tsuji,K., Yokoyama,Y. and Seino, Y. (1991) Assignment of Human Porphobilinogen Deaminase to 11q24.1-q24.2 by in situ Hybridization and Gene Dosage Studies. Cytogenet. Cell Genet. 57: 105-108.

Nelson D.L., Ledbetter S., Corbo,L., Victoria, M.F., Ramirez-Solis,R., Webster,T.D., Ledbetter D.H. and Caskey,C.T. (1989) Alu Polymerase Chain Reaction: A Method for Rapid Isolation of Human-Spcific Sequences from Complex DNA Sources. Proc.Natl. Acad. Sci. 86: 6686-6690.

Ngugen,C., Mattei, M.G., Goridis,C., Mattei,J.F. and Jordan, B.R. (1985) Localization of the Human N-CAM Gene to Chromosome 11 by in situ Hybridization with A Murine N-CAM cDNA Probe. (Abstract) Cytogenet. Cell Genet. 40: 713.

Oberle,I., Rousseau,F., heitz, D., Kretz,C., Devys,D., Hanauer,A., Boue,J., bertheas,M.F. and Mandel,J.L. (1991) Instability of A 500-Base Pair DNA Segment and Abnormal Methylation in Fragile X Syndrome. Science. 252: 1097-1102.

Olson,M., Hood,L., Cantor,C., and Botstein,D. (1989) A Common Language for Physical Mapping of the Human Genome. Science. 245: 1434-1435.

Orita, M., Suzuki, Y., Sekiya, T. and Yayashi, K. (1989) Rapid and Sensitive Detection of Point Mutations and DNA Polymorphisms Using the Polymerase Chain Reaction. Genomics. 5: 874-879.

Pritchard, C., Casher, D., Bull, L., Cox, D.R. and Myers, R.M. (1990) A Cloned DNA Segment from the Telomeric Region of Human Chromosome 4p Is Not Detectably Rearranged in Huntington Disease Patients. Proc. Natl. Acad. Sci. USA. 87: 7309-7313.

Reeders,S.T., Keith,T., Green,P., Germino,G.G., Barton,N.J., Lehmann,O.J., Brown,V.A., Phipps,P., Morgan,J., Bear,J.C. and Parfrey,P. (1988) Regional Localization of the Autosomal Dominant Polycystic Kidney Disease Locus. Genomics. 3:150-155.

Richard. E.J. and Ausubel, F.M. (1988) Isolation of A Higher Eukaryotic Telomere from Arabidopsis thaliana. Cell. 53: 127-136.

Rommens,J.M., iannuzzi,M.C., Kerem, B., Drumm,M.L., melmer,G., Dean,M., Rozmahel,R., Cole,J.L. Kennedy, D., Hidaka,N., Zsiga,M., Buchwald, M., Riordan,J.R., Tsui,L.-C. and Collins, F.S. (1989) Identification of the Cystic Fibrosis Gene:
Chromosome Walking and Jumping. Science. 245: 1059-1065.

Roussean-Merck,M.F., bernheim,A., Cherif,D., miglierina,R., Misrehi,M., loosfelt,H., Milgrom,E. and Berger,R. (1987) Localization of the Human Progesterone Receptor Gene (PGR) to Chromosome 11q22-q23. (Abstract) Cytogenet. Cell Genet. 46: 685.

Rouyer,F., Vergnaud,G., Johnsson,C., Levilliers,J., Petit,C. and Weissenbach,J. (1986) The Pseudoautosomal Region of the Human Sex Chromosomes. Cold Spring Harbor Symposia on Quantitive Biology. LI: 221-228.

Rowley, J.D., Diaz, M.O., Espinosa,R., Patel,Y.D., Melle,E.V., Ziemin S., Taillon-Miller, P., Lichter, P., Evans G.A., Kersey,J.H.,Ward,D.C., Domer,P.H. and Le Bean,M.M. (1990) Mapping Chromosome Band 11q23 in Human Acute Leukemia with Biotinylated Probes: Indentification of 11q23 Translocation Breakpoints with a Yeast Artificial Chromosome. Proc.Natl.Acad.Sci.USA. 87: 9358-9362.

Royle, N.J., Clarkson, R., Wong, Z. and Jeffrey, A.J. (1987) Preferential Localization of Hypervariabl Minisatellites Near Human Telomeres. Cytogenet Cell Genet. 46: 685-686.

Royer-Pokora, B., Kunkel,L.M., Monaco,A.P., Goff, S.C., Newberger,P.E., Baehner,R.L., Cole,F.S., Curnutte,J.T. and Orkin,S.H. (1986) Cloning the Gene for An Inherited Human Disorder-Chronic Granulomatous Disease-on the Basis of Its Chromosomal Location. Nature. 322: 32-38.

Saccone,S., De Sario A., Della Valle, G. and Bernardi, G. (1992) The Highest Gene Concentrations in the Human Genome Are in Telomeric Bands of Metaphase Chromosomes. Proc. Natl. Acad. Sci. USA. 89: 4913-4917.

Sambrook,J., Fritsch,E.F. and Maniatis, T. (1989) Molecular Cloning. Cold Spring Harbor Laboratory Press.

Sarkar,G., Kapelner,S., and Sommer,S.S. (1990) Formamide Can Dramatically Improve the Specificity of PCR. Nucl. Acids. Res. 18: 7465.

Selleri,L.,Hermanson,G.G.,Eubanks,T.H.,Lewis,K.A.,and Evans,G.A. (1990) Molecular Localization of the t(11;22) (q24;q12) Translocation of Ewing Sarcoma by Chromosomal in situ Suppression Hybridization. Proc.Natl.Acad.Sci.USA. 88: 887-891.

Shuldiner,A.,Nirula,A. and Roth,J. (1989) Hybrid DNA Artifact from PCR of Closely Related Target Sequences. Nucl.Acids. Res. 17: 4409.

Slightom,J.L., Theisen,T.W., Koop,B.F. and Goodman,M.(1987) Orangutan Fetal Globin Genes: Nucleotide Sequences Reveal Multiple Gene Conversions During Hominid Phylogeny. J. Biol.Chem. 262: 7472.

Smith,M., Smalley,S., Cantor,R., Pandolfo,M., Gomez,M.I., Baumann,R., Foldman,P., Yoshiyama,K., Nakamura,Y., Julier,C., Dumars,K., Haines,J., Trofatter,J., Spence,M.A., Weeks,D. and Conneally,M. (1990) Mapping of A Gene Determining Tuberous Sclerosis To Human chromosome 11q14-11q23. Genomics. 6: 105-114.

Speer,M.C. Yamaoka, L.H., Gilchrist,J.H., Gaskell,C.P., Stajich,J.M., Vance, J.M., Kazantsev,A., Lastra, A.A., Haynes, C.S., Beckmann, J.S., Cohen D., Weber, J.L., Roses,A.D. and Pericak-Vance,M.A. (1992) Confirmation of Genetic Heterogeneity in Limb-Girdle Muscular Dystrophy: Linkage of An Autosomal Dominant Form to Chromosome 5q. Am.J.Hum.Genet. 50: 1211-1217.

St. Clair., Blackwood,D., Muir,W., Carothers,A., Walker,M., Spowart, G., Gosden,C. and Evans,H.J. (1990) Association within A Family of A Balanced Autosomal Translocation with Major Mental Illness. Lancet. 336: 13-16.

Tautz,D. (1989) Hypervariability of Simple Sequences As A General Source for Polymorphic DNA Markers. Nucl. Acid.Res. 17: 6463-6471.

Tokino,T., Takahashi, E., Mori,M., Tanigami,A., Glaser,T., Park,J.W., Jones,C., Hori, T. and Nakamura,Y. (1991) Isolation and Mapping of 62 New Markers on Human Chromosome 11. Am. J. Hum. Genet. 48: 258-268.

Tunnacliffe,A. and McGuire,R.S.(1990) A Physical Linkage Group in Human Chromosome Band 11q23 Covering A Region Implicated in Leukocyte Neoplasia. Genomics. 8: 447-453.

Van der Ploeg, L.T., Liu, A.Y.C., and Brost, P. (1984) Structure of the Growing Telomeres of Trypanosomes. Cell. 36: 459-468.

Wallace, M.R., Marchuk,D.A., Andersen,L.B., Letcher,R., Odeh,H.M., Saulino,A.M., Foutain,J.W., Brereton,A., Nicholson,J., Mitchell,A.L., Brownstein, B.H. and Collins, F.S. (1991) Type 1 Neurofibromatosis Gene: Identification of A Large Transcript Disrupted in Three NF1 Patients. Science. 249: 181- 186.

Wang,Z. and Weber,J.L. (1992) Continuous Linkage Map of Human Chromosome 14 Short Tandem Repeat Polymorphisms. Genomics. 13: 536-536.

Weber,J.L.and May, P.E. (1989) Abundant Class of Human DNA Polymorphisms Which Can Be Typed Using the Polymerase Chain Reaction. Am.J.Hum.Genet.44: 388-396.

Weber,J.L. (1989) in The Polymerase Chain Reaction. Curr Comm in Molec Biol, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Weber,J.L.(1990) Human DNA Polymorphisms Based on Length Variations in Simple-Sequence Tandem Repeats. in Genome Analysis Volume 1: Genetic and Physical Mapping (David,K.E. and Tilghman S.M. ed.) Cold Spring Harbor Laboratory Press. Cold Spring Harbor.

Wei,S., Rocchi, M., Archidiacono,N., Sacchi,N., Romeo,G. and Gatti,R.A. (1990) Phisical Mapping of the Human Chromosome 11q23 Region Containing the Ataxia-Telangiectasia Locus. Cancer genet. Cytogenet. 46: 1-8.

Weissenbach,J., Levilliers,J., Petit,C., Rouyer F. and Simmler,M. (1987) Normal and Abnormal Interchanges between the Human X and Y Chromosomes. Development. 101 Supplement: 67-74.

Wilkie, A.O.M. Higgs, D.R., Raek, K.A., Buckle, V.J., Spurr, N.K., Fischel-Ghodsian. N., Cecherini, I., Brown. W.R.A. and Harris, P.C. (1991) Stable Length Polymorphism of Up To 260

Kb at the Tip of the Short Arm of Human Chromosome 16. Cell. 64: 595-606.

Wilkie, A. O. M. and Higgs, D.R. (1992) An Unususally Large (CA)n Repeat in the Region of Divergence between Subtelomeric Alleles of Human Chromosome 16p. Genomics. 13: 81-88.

Wijmenga C, Frants,R.R., Brouwer,O.F., Moerer,P., Weber,J.L.and Padberg,G.W. (1990) The Facioscapulohumeral Muscular Dystrophy Gene Maps To Chromosome 4. Lancet. 336: 651-653.

Zielenski,J., Markiewicz, D., Rininsland, F., Rommens, J. and Tsui,L.C. (1991) A Cluster of High Polymorphic Dinucleotide Repeats in Intron 17b of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Gene. Am. J. Hum.Genet. 49: 1256-1262.

Ziv,Y., Rotman,G., Frydman,M., Dagan,J.,Cohen,T., Foroud,T., Gatti,R.A. and Shiloh, Y. (1991) The ACT (Ataxia-Felangiectasia Complementation Group C) Locus Localizes to 11q22-q23. Genomics. 10: 373-375.

GENOMICS 10, 527–530 (1991)

Detection and Characterization of Additional DNA Polymorphisms in the Dopamine D2 Receptor Gene

XUE YA HAUGE,* DAVID K. GRANDY,‡ JAMES H. EUBANKS,§ GLEN A. EVANS,§ OLIVIER CIVELLI,‡ AND MICHAEL LITT*,†

*Departments of *Medical Genetics and †Biochemistry and Molecular Biology and ‡Vollum Institute for Advanced Biomedical Research, Oregon Health Sciences University, Portland, Oregon 97201; and §Molecular Genetics Laboratory, Salk Institute, La Jolla, California*

Received November 14, 1990; revised March 5, 1991

The gene encoding the dopamine D2 receptor (DRD2) has been suggested as a candidate gene for several mental disorders. We previously described the cloning and chromosomal mapping (to 11q22–q23) of a human DRD2 gene as well as its use for the detection of a two-allele *Taq*I RFLP with a minor allele frequency of 0.24, corresponding to a PIC of 0.30. Family linkage studies utilizing DRD2 would be facilitated if the PIC of the DRD2 locus were increased. To this end, we have used additional phage and cosmid clones in the vicinity of DRD2 to identify a new two-allele *Taq*I RFLP as well as a TG microsatellite polymorphism with a PIC of 0.62. We report localizations of the three polymorphisms on the restriction map of the DRD2 locus. The *Taq*I RFLPs are in apparent linkage equilibrium with the microsatellite, yielding a highly informative compound marker locus with a PIC of 0.76.

INTRODUCTION

It has been suggested that dopamine D2 receptors are involved in human psychoses, since many potent antipsychotic drugs have high affinities for these receptors (Seeman and Lee, 1975). Although several recent studies have failed to detect linkage between DRD2 and schizophrenia (Moises *et al.*, 1989), manic depression (Byerly *et al.*, 1989), or Tourette syndrome (Gelernter *et al.*, 1989), it remains possible that defects in DRD2 may underlie susceptibility to these or other diseases in some families.

Previously, we described the cloning and chromosomal mapping of a human DRD2 gene on 11q22–q23 (Grandy *et al.*, 1989a,b) as well as the detection at this locus of a two-allele *Taq*I RFLP with a minor allele frequency of 0.24, corresponding to a PIC of 0.30. This RFLP was detected with the genomic phage clone λhD2G1, which contains exon 8 and the 3′ untranslated portion of the DRD2 gene. Attempts to increase the PIC by screening 30 additional restriction enzymes with this probe failed to reveal additional polymorphisms. Since the testing of hypotheses of genetic linkage between DRD2 and human diseases would be facilitated by increasing the PIC at the DRD2 locus, we attempted to find additional polymorphisms within or close to the DRD2 gene. Because of the abundance, wide dispersion, and frequently high PICs of TG microsatellite polymorphisms in the human genome (Weber and May, 1989; Litt and Luty, 1989; Litt *et al.*, 1990), we screened several cosmids from the vicinity of DRD2 for TG microsatellites. We also used the clone λhD2G2, which contains coding exons 2–7, to screen for additional RFLPs at the DRD2 locus.

Here we report the discovery and characterization of two new polymorphisms at the DRD2 locus. We also present data describing studies of linkage disequilibria between marker loci within the DRD2 gene.

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. X54392.

MATERIALS AND METHODS

Cloning and DNA Sequencing

Construction of a cosmid library from the TG 5D1-1 cell line, which contains the q13–qter region of chromosome 11 as the only human genetic material, has been described (Evans and Lewis, 1989). This library was probed with a single-copy 1.6-kb *Bam*HI fragment subcloned from the DRD2 receptor locus. Cosmid c24,25, which hybridized with the DRD2 probe, also gave a signal when probed with the synthetic oligomer $(dC\text{-}dA)_{10}$ and was therefore selected for further study. *Sau*3A subclones of c24,25 were screened with the $(dC\text{-}dA)_{10}$ oligomer, and a positive subclone, 1:90, was selected for dideoxy sequencing. The relevant portion of the sequence of clone 1:90 (EMBL Accession No. X54392) is shown in Fig. 1.

0888-7543/91 $3.00

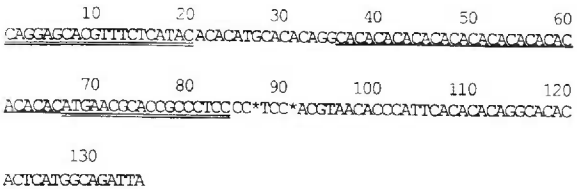


FIG. 1. Sequence of a 135-bp region of subclone 1:90 in the neighborhood of a $(CA)_{16}$ microsatellite repeat (positions 36–67, underlined). The sequences used to design PCR primers (positions 1–20 and 67–84) are double underlined. Asterisks at positions 88 and 91 indicate ambiguities in the sequence data.

The overlapping genomic clones λhD2G1 and λhD2G2 have been described (Grandy *et al.*, 1989b). A genomic clone containing exons 2–7 of the human D2 receptor gene has also been described by Dal Toso *et al.* (1989).

Genomic Southern Analysis

Separation of *Taq*I-digested DNA fragments, blotting, and hybridization to ^{32}P-labeled λhD2G2 in the presence of total human DNA as competitor were performed as previously described (Grandy *et al.*, 1989b).

Typing of Microsatellite Polymorphism

Using the sequence shown in Fig. 1, primers 509 (CAGGAGCACGTTTCTCATAC) and 419 (GGAGGGCGGTGCGGTCAT) were designed to amplify the microsatellite. We performed PCR in a total volume of 25 μl containing 100 ng of genomic DNA, 5 pmol of each primer (one of which was 5′-end-labeled with ^{32}P to a specific activity of ca. 10^6 dpm/pmol), 50 m*M* KCl, 10 m*M* Tris–Cl$^-$, pH 8.3, 1.5 m*M* $MgCl_2$, 200 μ*M* dATP, dTTP, dGTP, and dCTP, 0.6 unit of *Taq* polymerase (Perkin–Elmer/Cetus), and 0.01% gelatin. Amplification in an MJ Research thermal cycler was for 30 cycles; each cycle consisted of 30 s denaturation at 94°C, 30 s annealing at 58°C, and 30 s extension at 74°C. The final extension step was prolonged by 5 min. PCR products were resolved by electrophoresis on 6% denaturing polyacrylamide DNA sequencing gels and were detected by overnight autoradiography without an intensifying screen. Fragment sizes were measured relative to size standards consisting of 5′-end-labeled *Sau*3A fragments of pBR322 and/or DNA sequence ladders derived from a known sequence.

RESULTS AND DISCUSSION

λhD2G2 Reveals a TaqI RFLP

When λhD2G2 was used as a probe vs Southern blots of *Taq*I-digested human genomic DNA from unrelated individuals, it revealed an RFLP. Allelic fragments B1 and B2 had sizes of 4.6 and 4.1 kb, respectively. Constant bands at 10, 2.7, 2.6, and 1.4 kb were present in all individuals. Allele frequencies measured in 35 unrelated Caucasians were B1 = 0.16, B2 = 0.84. Codominant Mendelian inheritance was observed in two informative families with a total of 25 children (data not shown.)

Characterization of the Microsatellite Polymorphism

When primers 509 and 419 were used to amplify the TG/CA repeat, a VNTR polymorphism with four common alleles was revealed. Sizes of allelic fragments (nt) were C1 = 80, C2 = 82, C3 = 84, C4 = 86. Allele frequencies measured in 69 unrelated Caucasians were C1 = 0.16, C2 = 0.22, C3 = 0.47, C4 = 0.15. Codominant Mendelian inheritance was observed in three informative families with a total of 30 children. Figure 2 shows an autoradiogram of a gel used to type the TG microsatellite polymorphism in CEPH family 1332.

Localization of Polymorphisms

Figure 3 shows a partial *Bam*HI, *Bgl*II, and *Taq*I restriction map of the two overlapping genomic phage λhD2G1 and λhD2G2 and cos c24,25. Superimposed on the map are the locations of the seven D2 receptor coding exons, the positions of the two polymorphic *Taq*I sites (asterisks), and the TG microsatellite (underlined). The mapping of the polymorphic *Taq*I A site revealed by λhD2G1 was accomplished by *Taq*I digestion of cloned *Bam*HI fragments and genomic Southern blotting. The polymorphic *Taq*I B site was revealed with the λhD2G2 probe. Since *Taq*I digestion of this clone failed to produce fragments with sizes of either allele B1 (4.6 kb) or B2 (4.1 kb), we concluded that the polymorphic *Taq*I fragments were not wholly contained within the insert of this clone. This implied that the *Taq*I B site should be detectable with probes originating from one end or the other of the λhD2G2 insert. This expectation was confirmed when we found that a 3.7-kb *Bam*HI fragment of

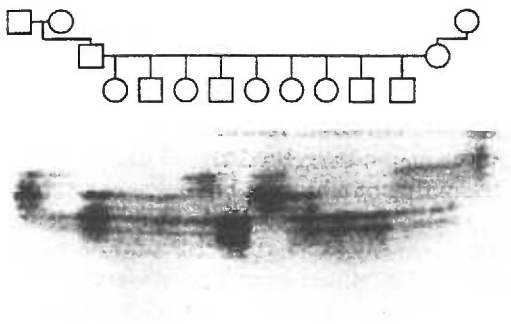


FIG. 2. Scoring of DRD2 microsatellite VNTR. Genomic DNAs from members of CEPH family 1332 were amplified using primers 419 and 509, as described in the text. The figure shows an autoradiogram of the acrylamide gel used to resolve the PCR products.

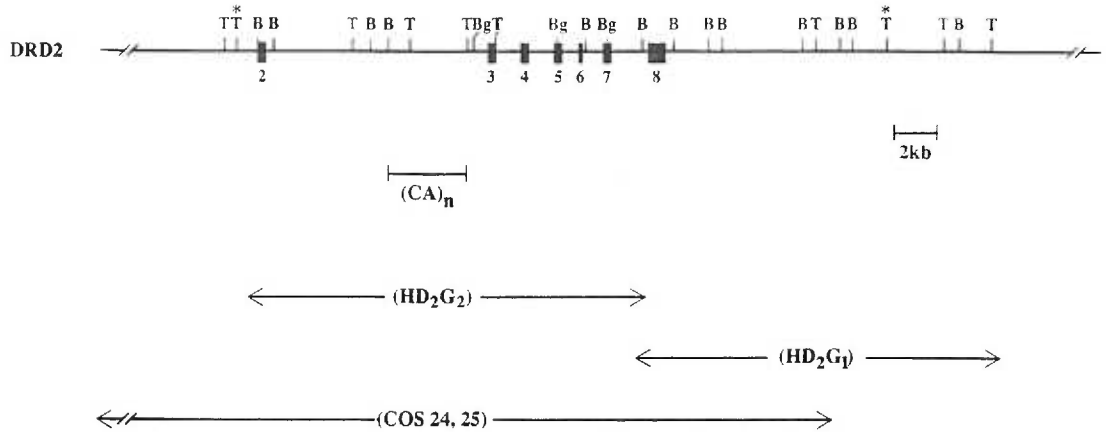


FIG. 3. Partial map of the human dopamine D2 receptor gene locus (DRD2). A partial restriction map is presented for the enzymes *Bam*HI (B), *Bgl*II (Bg), and *Taq*I (T). Polymorphic *Taq*I sites are indicated by asterisks and the region containing the microsatellite repeat is underlined. The regions corresponding to genomic phage λhD2G1 and λhD2G2 and cos c24,25 are also indicated.

λhD2G2, located near the 5′ end of the insert, hybridized to 4.6- and/or 4.1-kb *Taq*I fragments in human genomic DNA (data not shown). This indicated that the *Taq*I B site is located 5′ of the first coding exon of the D2 receptor gene. When a $(CA)_{10}$ oligomer was used to probe a Southern blot containing restriction digests of λhD2G2, it hybridized with a 3.5-kb *Bam*HI–*Bgl*II fragment, indicating localization of the microsatellite to the intron separating coding exons 2 and 3 (see Fig. 3).

Linkage Disequilibrium

Recently, it has been reported that there is association between the A1 allele of the *Taq*I A RFLP and severe alcoholism (Blum *et al.*, 1990), suggesting that a defect in or near DRD2 might predispose to alcoholism. In addition, it has been claimed that the allele frequencies at the *Taq*I A RFLP differ significantly between Tourette syndrome patients and controls (Comings *et al.*, 1990). We hoped that the microsatellite VNTR might be a useful tool for extending these types of studies. To investigate this possibility, we used samples from randomly ascertained unrelated Caucasians to measure linkage disequilibrium between pairs of marker loci within DRD2.

Tables 1 and 2 present data on pairwise linkage disequilibria for *Taq*I A and B and for *Taq*I A and microsatellite polymorphisms, respectively. Because the expected number of haplotypes is less than 5 for certain haplotypes in our dataset, the conclusions drawn from these data must be regarded as tentative. However, the data suggest strong linkage disequilibrium between the two *Taq*I RFLPs. In contrast, the *Taq*I A RFLP appears to be in linkage equilibrium with the microsatellite polymorphism. Eight haplotypes are observed with frequencies close to those expected for random association. Since the *Taq*I A and *Taq*I B sites appear to be strongly correlated, the above results suggest that the *Taq*I B site is also in equilibrium with the microsatellite polymorphism.

When we consider the localization of the polymorphic sites, it appears that the two polymorphic *Taq*I sites show a higher degree of disequilibrium with each other than either one does with the microsatellite locus in between them. Although this type of observation is unusual, similar observations have been reported previously. Barker *et al.* (1984) measured pairwise linkage disequilibria at the D11S12 locus and found that the *Msp*I and *Taq*I sites, which are the furthest separated pair of RFLPs at this locus, show stronger disequilibrium with each other than with other RFLPs located between them. Thompson *et al.* (1988) measured pairwise linkage disequilibria between closely linked markers in the AI-CIII apolipoprotein gene cluster and found, using conventional statistical tests, that *Msp*I and *Sst*I RFLPs were in stronger disequilibrium with each other than with a

TABLE 1

Observed and Expected Haplotypes for the *Taq*I A and B RFLPs in Unrelated Individuals

Loci				
A	B	Haplotype	*N*(obs)	*N*(exp)
1	1	1	5	0.94
1	2	2	7	11.1
2	1	3	0	4.1
2	2	4	52	48
Total			64	
$\chi^2 = 23.5$, $df = 1$, $P < 0.0001$				

Note. The χ^2 value characterizes the goodness-of-fit between observed and expected haplotype numbers and is followed by the degrees of freedom and the probability value, calculated according to the method of Hartl (11).

TABLE 2

Observed and Expected Haplotypes for the *TaqI* A and Microsatellite Polymorphisms (Locus C) in Unrelated Individuals

Loci				
A	C	Haplotype	*N* (obs)	*N* (exp)
1	1	1	1	1.7
1	2	2	1	2.7
1	3	3	9	6.3
1	4	4	2	2.3
2	1	5	7	6.3
2	2	6	12	10.3
2	3	7	21	23.7
2	4	8	9	8.7
Total			62	

$\chi^2 = 3.3$, $df = 3$, $P > 0.3$

*Pst*I RFLP located between them. The latter authors pointed out that failure to detect linkage disequilibrium using conventional statistical tests does not necessarily imply its absence, because very large sample sizes may be required to detect negative (i.e., repulsion phase) disequilibrium. This explanation may also apply to our failure to detect disequilibrium between the microsatellite and the *Taq*I RFLPs at the DRD2 locus. Alternatively, this lack of disequilibrium may be real and could be due to generation of new alleles at the microsatellite locus by a mechanism, such as replication slippage (Levinson and Gutman, 1987), which does not involve recombination.

Using haplotype frequencies calculated from the data in Table 2, we calculate a PIC (Botstein *et al.*, 1980) of 0.76, making DRD2 a strong candidate for inclusion on the "index map" of chromosome 11 (Roberts, 1990) and greatly facilitating its use in family linkage studies of alcoholism and other diseases.

ACKNOWLEDGMENTS

This work was supported by NIH Grants HG00022 and NS25370-02 (to M.L.) and NIMH Grant MH45614 (to O.C.). We thank Lynn Jorde for helpful discussions.

REFERENCES

1. Barker, D., Holm, T., and White, R. (1984). A locus on chromosome 11p with multiple restriction site polymorphisms. *Am. J. Hum. Genet.* **36:** 1159–1171.
2. Blum, K., Noble, E. P., Sheridan, P. J., Montgomery, A., Ritchie, T., Jadadeeswaran, P., Nogami, H., Briggs, A. H., and Cohn, J. B. (1990). Allelic association of human dopamine D2 receptor gene in alcoholism. *J. Am. Med. Assoc.* **263:** 2055–2060.
3. Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32:** 314–331.
4. Byerley, W., Mellon, C., Holik, J., Lubbers, A., Leppert, M., O'Connell, P., Reimherr, F., Grosser, B., Wender, P., Bunzow, J., Grandy, D., Civelli, O., Lalouel, J-M., White, R., and Litt, M. (1989). Molecular genetic studies using D2 dopamine receptor. Poster presented at American Psychiatric Association 142nd Annual Meeting, May 6–11, 1989, San Francisco.
5. Comings, D. E., Muhleman, D., Dietz, G., Shahbahrami, B., Tast, D., and Kovacs, B. W. (1990). *Am. J. Hum. Genet.* **47:** A52.
6. Dal Toso, R., Sommer, B., Ewert, M., Herb, A., Pritchett, D. B., Bach, A., Shivers, B. D., and Seeburg, P. H. (1989). The dopamine D2 receptor: Two molecular forms generated by alternative splicing. *EMBO J.* **8:** 4025–4034.
7. Evans, G. A., and Lewis, K. A. (1989). Physical mapping of complex genomes by cosmid multiplex analysis. *Proc. Natl. Acad. Sci. USA* **86:** 5030–5034.
8. Gelernter, J., Pakstis, A., Chappell, P., Kurlan, R., Grandy, D. K., Bunzow, J., Retief, A. E., Litt, M., Civelli, O., and Kidd, K. K. (1989). Tourette's syndrome is not linked to D2 receptor. Poster presented at American Psychiatric Association 142nd Annual Meeting, May 6–11, 1989, San Francisco.
9. Grandy, D. K., Marchionni, M., Makam, H., Stofko, R. E., Alfano, M., Frothingham, L., Fischer, J. B., Burke-Howie, K. J., Bunzow, J. R., Server, A. C., Magenis, R. E., and Civelli, O. (1989a). Cloning of the cDNA and gene for a human D2 dopamine receptor. *Proc. Natl. Acad. Sci. USA* **86:** 9762–9766.
10. Grandy, D. K., Litt, M., Allen, L., Bunzow, J. R., Marchionni, M., Makam, H., Reed, L., Magenis, R. E., and Civelli, O. (1989b). The human dopamine receptor gene is located on chromosome 11 at q22–q23 and identifies a TaqI RFLP. *Am. J. Hum. Genet.* **45:** 778–785.
11. Hartl, D. L. (1980). "Principles of Population Genetics," pp. 110–111, Sinauer, Sunderland, MA.
12. Levinson, G., and Gutman, G. A. (1987). Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4:** 203–221.
13. Litt, M., and Luty, J. A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **44:** 397–401.
14. Litt, M., Guo, Z., Luo, X. Y., Luty, J. A., and Sharma, V. (1990). Chromosomal distribution of polymorphic TG microsatellites. *Am. J. Hum. Genet.* **46:** A189.
15. Moises, H. W., Gelertner, J., Giuffra, L. A., Kidd, J. R. L., Pakstis, A. J., Bunzow, J. R., Grandy, D. K., Sjogren, B., Wetterberg, L., Kennedy, J., Litt, M., Civelli, O., Kidd, K. K., and Cavalli-Sforza, L. L. (1989). Exclusion of the D2 dopamine receptor gene as candidate gene for schizophrenia in a large pedigree from Sweden. First World Conference on Psychiatric Genetics, Cambridge, UK, August 3–5, 1989.
16. Roberts, L. (1990). The genetic map is back on track after delays. *Science* **248:** 805.
17. Seeman, P., and Lee, T. (1975). Antipsychotic drugs: Direct correlation between clinical potency and presynaptic action on dopamine neurons. *Science* **188:** 1217–1219.
18. Thompson, E. A., Deeb, S., Walker, D., and Motulsky, A. G. (1988). The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. *Am. J. Hum. Genet.* **42:** 113–124.
19. Weber, J. L., and May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44:** 388–396.

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

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.

Nucleic Acids Research 18: 5920 (1990)

Dinucleotide repeat polymorphism at the D11S420 locus

X.Y.Luo, G.A.Evans[1] and M.Litt*
Oregon Health Sciences University, Portland, OR 97201 and [1]Salk Institute, La Jolla, CA 92138, USA

Source and Description of Clone: Cosmid 8.5 was from a human chromosome 11q-specific library (1). p8.5−3, a subclone of this cosmid, was sequenced and the sequences flanking a $(GT)_{18}$ repeat (EMBL accession number X52290) were used to design PCR primers.

PCR Primers:
(#506)-5′-AGTTACACCGGTTCTGCAGA-3′
(#507)-5′-GATTAATGATAGTGCTATCC-3′

Polymorphism: Allelic fragments were detected on DNA sequencing gels. Lengths (nt) were: A1=188, A2=190, A3=192, A4=194, A5=196, A6=208.

Frequencies: From 33 unrelated European Caucasians, A1=.39, A2=.36, A3=.063, A4=.063, A5=.063, A6=.063. The PIC calculated from these frequencies is 0.66.

Chromosomal Localization: 11q23.3-q24, by fluorescent in situ hybridization (1).

Mendelian Inheritance: Mendelian inheritance was observed in 2 informative CEPH families with a total of 22 children.

PCR Conditions: PCR was performed 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. Primer #506 was 5′ end labeled with ^{32}P. Amplification was for 30 cycles with denaturation at 94°C for 1 min, annealing at 49°C for 1 min and extension at 72°C for 30 sec. Amplified products were resolved on DNA sequencing gels and detected by autoradiography.

Acknowledgements: This work was supported by Grants GM-32500 and HD17449 from the National Institutes of Health.

Reference: 1)Lichter,P., Tang,C.C., Call,K., Hermanson,G., Evans,G.A., Housman,D. and Ward,D.C. (1990) *Science* **24**, 64−69.

Nucleic Acids Research 18: 7470 (1990)

Dinucleotide repeat polymorphism at the D11S490 locus

X.Y.Luo, G.A.Evans[1] and M.Litt*
Oregon Health Sciences University, Portland, OR 97201
and [1]Salk Institute, La Jolla, CA 92138, USA

Source and Description of Clone: Cosmid 1,16 was from a human chromosome 11q-specific library (1). p1,16−2, a subclone of this cosmid, was sequenced and the sequences flanking a $(GT)_{14}$ repeat (EMBL accession no. X56514) were used to design PCR primers.

PCR Primers:
(#22314)-5′-CACAAACATTGGCGCAT-3′
(#41919)-5′-TTCTGGGTCACGGTGCTTCA-3′

Polymorphism: Allelic fragments were detected by autoradiography after separation on denaturing polyacrylamide gels. Lengths (nt) were: A1 = 167, A2 = 165, A3 = 163, A4 = 161, A5 = 159, A6 = 157, A7 = 155, A8 = 153, A9 = 147.

Frequencies: from 50 unrelated European Caucasians. A1 = .01, A2 = .07, A3 = .12, A4 = .07, A5 = .09, A6 = .43, A7 = .01, A8 = .02, A9 = .18. The PIC calculated from these frequencies is 0.72.

Chromosomal Localization: 11q23.3, by fluorescence in-situ hybridization (1).

Mendelian Inheritance: Mendelian inheritance was observed in 2 informative CEPH families with a total of 17 children.

PCR Conditions: PCR was performed according to reference (2), using 5′ end labeled primer #22314 and an annealing temperature of 54°, 35 cycles. Amplified products are resolved on denaturing acrylamide gels of this type used for DNA sequencing and detected by autoradiography.

Acknowledgements: This work was supported by Grants GM-32500 and HD 17449 from the National Institutes of Health.

References: 1) Lichter,P., Tang,C.C., Call,K., Hermanson,G., Evans,G.A., Housman,D. and Ward,D.C. (1990) *Science* 24, 64−69. 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.

Nucleic Acids Research 19: 4790 (1991)

Dinucleotide repeat polymorphism at the D11S527 locus

D.L.Browne, J.Gault, M.B.Thompson, X.Y.Hauge, G.A.Evans[1] and M.Litt*
Oregon Health Sciences University, Portland, OR 97201 and [1]Salk Institute, La Jolla, CA 92138, USA

Source and Description of Clone: Cosmid 8G11, also known as c23,23 or cosmid ZD5 was from a human chromosome 11q specific library (1). A subclone of this cosmid (pJG4) was isolated by hybridization to a $(CA)_{15}$ oligonucleotide and partially sequenced. The sequences (EMBL accession number: X60166) flanking the repeat $(GT)_{17}(CTGT)_6$ were used to design PCR primers.

PCR Primers:
JG4-A = 5′-GCCCCTCTACTTGTCTGGAG-3′
JG4-C = 5′-ATGCGGCTCCAAGACAAGTTC-3′

Polymorphism: Allelic fragments were detected on DNA sequencing gels. Lengths (nt) are: A1 = 166, A2 = 164, A3 = 162, A4 = 160, A5 = 158, A6 = 156, A7 = 154, A8 = 152, A9 = 150, A10 = 148, A11 = 146, A12 = 142. Alleles in four CEPH parents were as follows: 134101: A4,A9; 134102: A10,A10; 141801: A4,A9; 141802: A3,A11.

Frequencies: from 77 unrelated CEPH parents: A1 = 0.013, A2 = 0.11, A3 = 0.097, A4 = 0.156, A5 = 0.162, A6 = 0.058, A7 = 0.136, A8 = 0.019, A9 = 0.091, A10 = 0.078, A11 = 0.071, A12 = 0.006. The PIC calculated from these frequencies is 0.88.

Mendelian Inheritance: Mendelian inheritance was observed in 10 informative CEPH families with a total of 79 children.

Chromosomal Localization: Cosmid ZD5 was mapped to 11q13.5 by fluorescent in situ hybridization (1). Three-point linkage analysis of 10 informative CEPH families suggests the order cen-D11S97-D11S527-D11S388-qter, which is consistent with the in situ hybridization data. Odds against inversion of D11S527 and D11S97 are 100:1; odds against inversion of D11S527 andD11S388 are 3800:1.

PCR Conditions: PCR was performed according to (2) except that the products were not radioactively labelled. Annealing temperature was 60°C. 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.

Acknowledgements: This work was supported by NIH Grant HG00022 and by a grant from the Retinitis Pigmentosa Foundation Fighting Blindness.

References: 1) Lichter,P., Tang,C.C., Call,K., Hermanson,G., Evans,G.A., Housman,D. and Ward,D.C.(1990) *Science* **24**, 64–69. 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.

Nucleic Acids Research 19: 1964 (1991)

Dinucleotide repeat polymorphism at the D11S528 locus

X.Y.Hauge, G.A.Evans[1] and M.Litt*
Oregon Health Sciences University, Portland, OR 97201
and [1]Salk Institute, La Jolla, CA 92138, USA

Source and Description of Clone: Cosmid 13,1 was from a human chromosome 11q-specific library (1). p13,1−1 a subclone of this cosmid, was sequenced and the sequences flanking a $(AC)_{17}$ repeat (EMBL accession number X56995) were used to design PCR primers.

PCR Primers:
(#42026)-5′-AATGGTGTCCCCACACATGT-3′
(#42027)-5′-TCCTACCTACCGAGCTTAAA-3′

Polymorphism: Allelic fragments were detected on DNA sequencing gels. Lengths (nt) were: A1 = 91, A2 = 89, A3 = 87, A4 = 83, A5 = 81, A6 = 79, A7 = 77, A8 = 75, A9 = 73.

Frequencies: From 28 unrelated European Caucasians.

A1 = .02	A2 = .02
A3 = .02	A4 = .05
A5 = .02	A6 = .04
A7 = .23	A8 = .05
A9 = .55	

The PIC calculated from these frequencies is 0.60.

Chromosomal Localization: Linkage analysis with THY1 (localized to 11q22.3-q23 (2)) in 5 informative CEPH families gave a maximum lod score of 3.2 at θ = .064.

Mendelian Inheritance: Mendelian inheritance was observed in 5 informative CEPH families with a total 39 children.

PCR Conditions: PCR was performed in a total volume of 25 µl containing: 50 ng genomic DNA, 5 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 of Taq polymerase (Perkin-Elmer/Cetus) and 0.01% gelatin. Primer #42026 was 5′ end labeled with ^{32}P. Amplification was for 35 cycles with denaturation at 94°C for 1 min, annealing at 52°C for 1 min and extension at 74°C for 1 min. Amplified products were resolved on DNA sequencing gels and detected by autoradiography.

Acknowledgements: This work was supported by Grant HG-00022 from the National Institutes of Health.

References: 1)Lichter,P., Tang,C.C., Call,K., Hermanson,G., Evans,G.A., Housman,D. and Ward,D.C. (1990) *Science* 24, 64−69. 2)Junien,C. and McBride,O.W. (1989) *Cytogenet Cell Genet* 51, 226−258.

Nucleic Acids Research 19: 4308 (1991)

Dinucleotide repeat polymorphism at the D11S534 locus

X.Y.Hauge, G.A.Evans[1] and M.Litt*
Oregon Health Sciences University, Portland, OR 97201
and [1]Salk Institute, La Jolla, CA 92138, USA

Source and Description of Clone: Cosmid 30,1 was from a human chromosome 11q-specific library (1). p30,1−1, a Sau 3A subclone of this cosmid in the vector pTZ18u, was partially sequenced (EMBL accession no. X59147) and the sequences flanking a $(AC)_{20}$ repeat were used to design PCR primers.

PCR Primers:
(# 91249) −5′-ATATGGAAACTCTCCGTACT-3′
(# 92005) −5′-GCAACCATGGAGAGTCTGGA-3′

Polymorphism: Allelic fragments were detected on 6% denaturing polyacrylamide sequencing gels. Lengths (nt) were: A1 = 244, A2 = 242, A3 = 240, A4 = 238, A5 = 236, A6 = 234, A7 = 232, A8 = 230, A9 = 228. The parents of CEPH family 1332 had the following genotypes: 133201-A5, A6; 133202-A3, A8.

Frequencies: From 58 unrelated European Caucasians.
A1 = .008 A2 = .05
A3 = .15 A4 = .34
A5 = .10 A6 = .28
A7 = .03 A8 = .03
A9 = .012
The PIC calculated from these frequencies is 0.74.

Chromosomal Localization: Linkage analysis in 9 CEPH families with the MspI RFLP identified by the probe pHBI59 (D11S146 gave a maximum LOD score of 11.88 at theta = 0. The locus D11S146 has been mapped to chromosome 11q13 (2).

Mendelian Inheritance: Mendelian inheritance was observed in 9 informative CEPH families with a total of 70 children.

PCR Conditions: PCR was performed in a total volume of 12.5 μl containing 25 ng genomic DNA, 2.5 pmole of each primer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 50 mM KCl, 5 mM Tris Cl pH 8.3, 0.3 units of Taq polymerase (Perkin- Elmer/Cetus) and 0.01% gelatin. Primer # 91249 was 5′ end labeled with ^{32}P. Amplification was for 30 cycles with denaturation at 94°C for 1 min, annealing at 55°C for 1 min and extension at 74°C for 1 min. Amplified products were resolved on DNA sequencing gels and detected by autoradiography.

Acknowledgement: This work was supported by NIH Grant HG00022 and by a grant from the Retinitis Pigmentosa Foundation Fighting Blindness.

References: 1) Lichter,P., Tang,C.C., Call,K., Hermanson,G., Evans,G.A., Housman,D. and Ward,D.C. (1990) *Science* 24 64−69. 2) Julier,C., Nakamura,Y., Lathrop,M., O'Connell,P., Leppert,M., Litt,M., Mohandas,T., Lalouel,J. and White,R. (1990) *Genomics* 7, 335−345.

Cold Spring Harbor Meeting on Human Genome Mapping and Sequencing May, 1992.

Abstract

INDEX MAP OF THE LONG ARM OF HUMAN CHROMOSOME 11
Xueya Hauge, David Browne, and Michael Litt. Oregon Health Sciences University, Portland, OR.

A short-term goal of the US Human Genome Initiative is to construct "Index Maps" of all human chromosomes.These maps will consist of markers with minimum heterozygosities of 70 % and with no gaps exceeding 10-15 cM. We have accepted responsibility for the index map of chromosome 11. Here, we describe a partial index map consisting of a continuous group of markers extending from D11S288/D11S436 (localized to 11p12-p11.2) to D11S490 (localized to 11q23.3). The table below lists the markers in the order pter-qter and shows the sex-averaged pairwise recombination fractions between adjacent markers.

Marker	Heterozygosity	Marker type	Source of data
D11S436/ D11S288	0.74	RFLP haplotype	CEPH v5
	0.05		
D11S97	0.77	VNTR	CEPH v5
	0.06		
D11S533	0.88	hexanucleotide repeat	Litt lab
	0.05		
D11S527	0.88	CA repeat	Litt lab
	0.15		
D11S388	0.88	RFLP haplotype	CEPH v5
	0.12		
D11S35	0.86	CA repeat	Litt lab
	0.12		
DRD2	0.76	CA repeat/ RFLP haplotype	Litt lab
	0.03		
D11S144/ D11S351	0.79	RFLP haplotype	CEPH v5
	0.05		
D11S490	0.72	CA repeat	Litt lab
	ca. 0.40		
D11S127	0.71	VNTR	CEPH v5

The major gap remaining in this map is that between D11S490 (localized to 11q23.3) and D11S127 (localized to 11q24-q25.) Also,

isolation of a marker locus physically linked to the 11q telomere will be required in order to complete the map.

1/16-2 DNA sequence

```
          10          20          30          40          50
CAGAGTGTAG TGAAGTGTGA CACGAGGCAT TTCTTTTCTA GCTCCACCAT

          60          70          80          90         100
CCTGTGGCCT GGAAACAAAG ACTCATTCTG GTCACGGTG CTTCAGGGAG
                                                        **
         110         120         130         140         150
CTCTGCAGTG AGAGTGTGTG TGTGTGTGTG TGTGTGTGTG TGCGTGTACC
* *
         160         170         180         190         200
TGTGTGTTGA CTTGTGGCAG TGTGTTGTGG GTTGTTGTTT TTTTCCTAAA

         210         220         230         240         250
TGGCTTAATG CGCCAATGTT TGTGGAGCAT CCTTGGAATC AAAGTGTCTT

         260         270         280         290         300
TGGCGTAATT ATTGTTAATT GCGCAGACTA TTCTATAAAC TCATCCAATT

         310         320         330         340         350
AAGAGAAAGT TATTAACTTT ATAAAATATC GCAGGCATCC CACAGTGCCG

         360         370         380
TGTGTAGATG GAATAACAGT TTGAGAGACC
```

PCR primers are underlined.
* Alu recognition site.