THE HUMAN GENE MAP NEAR THE FRAGILE X

by Graeme Kemble Suthers.

This thesis is submitted as the complete requirement for the degree of Doctor of Philosophy within the Department of Paediatrics, Faculty of Medicine, at The University of Adelaide, South Australia.

August 1990.

Awarded 2.1.91

First of Two Volumes.

for Jenny,

... and Andrew and Nicholas.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University and, to the best of the candidate's knowledge and belief, contains no material previously published or written by another person, except where due reference is made in the text. Except where explicitly indicated in the text, the work described in this thesis was performed by the candidate or under the candidate's supervision. If this thesis is accepted for the award of the degree, the candidate consents to it being made available for photocopying and loan if applicable.

Graeme Suthers B.Sc.(Med.)(Hons I)(Sydney), M.B.B.S. (Hons I)(Sydney), F.R.A.C.P.

TABLE OF CONTENTS

Volume One

Volume Two

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The fragile X syndrome is the most common cause of familial mental retardation. It is characterised by mental retardation, subtle dysmorphic features, and a fragile site at Xq27.3. Segregation studies have demonstrated incomplete penetrance in males and females.

It was not known whether males with X-linked mental retardation who were fragile X negative had a disorder that was allelic to the fragile X syndrome. The locus responsible for non-specific X-linked mental retardation in a large pedigree was mapped to Xp11 by linkage analysis. This locus (MRX1) is distant from the fragile X locus (FRAXA) which is located at Xq27.3. The fact that MRX1 is not allelic to FRAXA indicates that non-specific X-linked mental retardation is genetically heterogeneous. The 98% confidence interval for MRX1 location was estimated by using a novel resampling strategy to be 0 to 9 centiMorgan (cM) distal to DXS14.

The development of a precise genetic map near FRAXA has been hampered by a lack of closely linked polymorphic loci. The closest restriction fragment length polymorphisms (RFLPs) to FRAXA lay 5 cM proximal (DXS369) and 3 cM distal (DXS304) to FRAXA. The established order of loci near FRAXA was cen-F9-DXS105-DXS98-DXS369-FRAXA-DXS304-DXS374-DXS52-qter.

A panel of 14 cell lines with X chromosome translocation or deletion breakpoints near FRAXA was assembled. The locations of the breakpoints were defined with 14 established probes. Seven of the cell lines had breakpoints between DXS369 and DXS304, and it was not possible to define further the locations of the X chromosome breakpoints in relation to FRAXA. One of these cell lines was derived from a female with Hunter syndrome (MPS II; iduronate-2-sulfatase [IDS] deficiency) due an X;autosome translocation, thus localizing IDS to between DXS369 and DXS304.

The panel of cell lines was used to localize 18 new DNA probes in this region. One probe was an IDS cDNA clone; the remainder were anonymous DNA fragments. Eight of the probes detected loci near FRAXA. The X chromosome breakpoints, the new probes, and IDS were all localized in relation to each other and to FRAXA. The order of probes and loci near FRAXA was:

cen-DXS369,VK24-VK47-VK23-VK16,FRAXA-VK21-VK18-IDS-VK37-DXS304-qter.

RFLPs were detected by the probes VK21 and VK23. RFLPs were also detected at IDS using the IDS cDNA clone. RFLPs were not detected with the probes VK16 or VK18. The RFLPs were mapped in normal pedigrees using the LINKAGE package of computer programs and programs written by the candidate. The following order of loci and recombination fractions were obtained:

6% 4% 12% 0% 6% 0% 1% 12%

F9--DXS105--DXS98---DXS369--VK23--VK21--IDS--DXS304----DXS52

This genetic map was used as the basis for a multipoint linkage study of the fragile X syndrome. 35 Australian pedigrees were genotyped for the three new RFLPs and other nearby polymorphisms. Genotypings were obtained from a further 77 pedigrees as part of an international collaborative project. On multipoint linkage analysis of these data, FRAXA was located 2 cM proximal to VK21. These results define a new diagnostic strategy for DNA studies of fragile X families. A combination of five probes and three restriction endonucleases will identify an RFLP within 4 cM of FRAXA in 94% of women.

The genetic maps derived from the normal families and from the fragile X families were compared. Contrary to previous reports, there was no difference in these genetic maps between or within the populations.

ACKNOWLEDGEMENTS

This project was carried out in the Cytogenetics Unit of the Department of Histopathology at the Adelaide Children's Hospital in the period 1987-1990. The Unit has since become the Department of Cytogenetics and Molecular Genetics within the Hospital.

It has been a great personal and professional pleasure to work at the Hospital. I would particularly like to thank Grant Sutherland, Eric Haan, John Mulley, Val Hyland, and David Callen, who introduced me to the intricacies of molecular genetics and linkage analysis. Grant was the model supervisor in every way.

At various times the following people have kept me sane, helped me out, fed me marzipan cookies, forgiven me for deleting their files, and generally made the last three years a most stimulating and enjoyable time: Agi Gedeon, Helen Kozman, Liz Baker, Rob Richards, Julie Nancarrow, Jill McCure, Kathy Friend, Don Perry, Athalie Nation, Norah Burton, Joan Westley, Sue White, Hilary Philips, Sharon Lane, Paul Nelson, Kirk Fernandez, and Lyndall Bates.

Most of the DNA methods used in this project were established in this laboratory by John Mulley, Antonio Fratini, Ieva Stupans, Arnie Puspurs, and Val Hyland. Many of the cell

lines and DNA probes used in this study were provided by other research groups (listed in Chapter 2). The following people kindly provided DNA probes prior to publication, K.E. Davies (1A1.1), N. Dahl (U6.2), B.A. Oostra (RN1), P.J. Wilson and C.P. Morris (pc2S15). DNA samples and genotypes for the forty CEPH families (and associated software) were provided by CEPH. Jurg Ott of Columbia University, New York, kindly provided the linkage programs, and freely (and very promptly) gave advice. Sue Wilson (A.N.U.) patiently taught me statistics. The librarians at the Hospital were very helpful.

The Professor of Paediatrics, Professor George Maxwell, and his successor, Professor Don Roberton, were very supportive. Professor Roberton critically reviewed a draft of this thesis. I would also like to acknowledge my gracious hosts, Dr Rod Carter (Director of the Department of Histopathology) and his successor, Dr Anthony Bourne. My interest in this field was stimulated during a very enjoyable stint in 1986 with Associate Professor Gillian Turner at the Prince of Wales Children's Hospital, Randwick NSW.

The candidate was supported by the National Health and Medical Research Foundation of Australia and The University of Sydney Post-graduate Committee in Medicine.

Note regarding publications.

Much of the material presented in this thesis has been published. These published articles are indicated when cited, and copies of the papers can be found in Appendix D.

Chapter 1

LITERATURE REVIEW.

PREFACE

The fragile X syndrome is the most common cause of familial mental retardation. It is an X-linked disorder with very unusual clinical and cytogenetic features. Nothing is known of the pathogenesis of the fragile X syndrome and little progress has been made towards isolating the fragile X mutation.

In this literature review the place of the fragile X syndrome in the history of mental retardation is discussed. The clinical and cytogenetic features of the syndrome and the results of segregation studies are then presented. Those studies clarified the major questions regarding the fragile X syndrome but essentially left them unanswered. A different approach was required to begin addressing the molecular pathology of this disorder.

The evolving human gene map has provided the necessary framework for approaching the fragile X mutation. The development of the human gene map will be reviewed, with particular reference to the role of genetic linkage studies in localizing genes of unknown function. The current status of the human gene map near the fragile X locus will be described. The review will conclude with a description of the rationale for the research presented in the remainder of this thesis.

INTRODUCTION

In 1930 the Medical Research Council of the United Kingdom supported an exhaustive study by Lionel Penrose of the causes of mental retardation at a large institution in Colchester. This was a critical time in the history of the study of human intelligence. The hereditarian theory of intelligence had come into disrepute in academic circles, partly because of the racial prejudice of some advocates of the Eugenic movement, and partly because intelligence was recognised to be a graded character that did not easily fit into a Mendelian mould (Kevles 1985, p.129). In the preface to Penrose's report, the Council acknowledged that political and social factors, as much as scientific ones, had delayed the study of mental retardation (Penrose 1938, p.1).

> *"Mental deficiency is a social problem of major importance... It is only in recent years, however, that students of social problems have come to recognize the importance and significance of mental deficiency. The reasons for this are obvious. Hitherto the Mental Health Service has been regarded as the infant among the general public health services; ... The importance of mental deficiency has been masked also by the fact that, administratively, the mental defective individual was simply one of a large number of destitute persons for whom the Poor Law authorities had to cater...The lack of definition of the problems to be faced, and the further practical difficulty that the methods of study of mental disorder are limited, ...have frightened away many investigators from this field of inquiry."*

By the time of Penrose's report there had been a number of surveys showing an excess of males in mentally retarded populations (Rosanoff 1931; others cited by Penrose 1938,

p.12 and by Lehrke 1974). Rosanoff had concluded that *"gene factors of intelligence may in some cases be carried in the X-chromosomes..."*. In view of the difficulties of grappling with the biological basis of mental retardation it is intriguing that this suggestion was not pursued. There are a number of reasons why this was so.

Firstly, there was considerable concern over the influence of the eugenic movement in public debate and legislation (Kevles 1985, p.147), and the environment was acknowledged to have a major impact on measured intelligence (Kevles 1985, p.141). Secondly, the discrepant sex-ratio was attributed to ascertainment bias due to society having different expectations of men versus women (Penrose 1938, p.2) and institutions having sex-biased admission criteria (Anastasi 1972). Thirdly, the data sometimes failed to support the theory of specific X-linked loci responsible for intelligence. In Penrose's study he noted a 25% excess of males but was unable to prove that this was due to sex-linked genes (Penrose 1938, p.66)[1]. Finally, the male excess was possibly attributed to sociological rather than genetic factors because many of the retarded men lacked any other evidence of a genetic defect - they had no phenotypic or

---

1 A sophisticated re-analysis of Penrose's data some 40 years later tentatively concluded that X-linked loci accounted for perhaps 5% of the cases studied (Morton et al. 1977).

biochemical abnormalities other than mental retardation (Penrose 1938, p.55).

However, if genes necessary for normal intellectual function were located on the X chromosome, one could expect that allelic variation at these loci would be more evident in males because they are hemizygous for X-linked loci. Men would be over-represented in retarded populations and also be over-represented in highly intelligent populations. Lehrke (1972a, 1972b, 1974) argued that the observed excess of both retarded and gifted males indicated that there are major X-linked loci responsible for intellect. The absence of physical or biochemical stigmata associated with either extreme of intellect indicated that the primary function of the X-linked loci related to brain function, and that "non-specific" X-linked mental retardation was not secondary to a more general error of metabolism. On the basis of population data Lehrke estimated that X-linked loci accounted for 25-50% of all mental retardation.

Lehrke's hypothesis relied heavily on data from the literature and attracted strong criticism regarding the suitability of his secondary sources (Anastasi 1972; Nance & Engel 1972). Nonetheless, he neatly encapsulated the potential biological and clinical significance of non-specific X-linked mental retardation: there are X-linked loci specifically associated with intellectual function; allelic

variation at these loci accounts, in part, for the range of intellect observed in human populations; deleterious mutations at these loci are common; and, as these loci are brain-specific, there are few phenotypic clues to differentiate the loci or the mutations.

At about this time Turner et al. (1971, 1972) and Davison (1973) documented that X-linked mental retardation was common in mentally retarded populations, and accounted for approximately 20% of mental retardation in males (Turner & Turner 1974). Gillian Turner went further and noted that many of the males with X-linked mental retardation were of normal appearance, and that "to be normal looking in a moderately mentally retarded population was relatively abnormal" (Turner 1983). Population surveys suggested the frequency of males with moderate X-linked mental retardation was 0.6 per 1,000 (Turner & Turner 1974), and of males with any degree of X-linked mental retardation was 1.8 per 1,000 (Herbst & Miller 1980).

It was indirectly estimated that non-specific X-linked mental retardation could be due to 7-19 different X-linked loci (Morton et al. 1977; Herbst & Miller 1980). These estimates of the number of brain-specific X-linked loci were critically dependent on two assumptions. The first was that the mutation rates at these loci were similar to those documented for other X-linked conditions such as Duchenne muscular dystrophy

or hemophilia A. Secondly, it was assumed that the hemizygous males had zero reproductive fitness while heterozygous females had normal reproductive fitness. It is now apparent that these assumptions are not true, at least in relation to the fragile X syndrome, and the number of X-linked brain-specific loci is still open to question.

In the absence of any phenotypic classification it was impossible to specifically analyze different X-linked loci. It was therefore of interest when the association of non-specific X-linked mental retardation and macro-orchidism was noted (Turner et al. 1975). Even more intriguing was the identification of a fragile site at the distal end of the long arm of the X-chromosome in some males with non-specific X-linked mental retardation (Lubs 1969; Giraud et al. 1976; Harvey et al. 1977; Sutherland 1977a). The first description of the fragile site pre-dated Lehrke's work but it was not until Sutherland described specific folic acid deficient culture media (Sutherland 1977b) that the fragile site could be reproducibly observed in a number of laboratories. Turner studied 16 pedigrees with non-specific X-linked mental retardation and found that in six pedigrees the affected males had macro-orchidism and expressed the fragile site (Turner et al. 1978, 1980). Subsequent studies of non-specific X-linked mental retardation indicated that macro-orchidism could occur in the absence of fragile site expression (Herbst et al. 1981; Fishburn et al. 1983).

The fragile X syndrome may be defined as consisting of X-linked mental retardation with expression of a folate-sensitive fragile site at Xq27 (the fragile X) in the absence of other phenotypic or biochemical abnormalities (Kaiser-McCaw et al. 1980). On this basis, the fragile X syndrome accounts for approximately 5% of all mental retardation and at least 25% of X-linked mental retardation (Sutherland & Hecht 1985, p.111). The first pedigree described with X-linked mental retardation and subsequently shown to have the fragile X syndrome was that reported by Martin and Bell(1943); the fragile X syndrome is occasionally referred to eponymously as the Martin-Bell syndrome.

THE FRAGILE X SYNDROME - Clinical Features

The main ascertainment criterion for identifying people with the fragile X syndrome has been mental retardation. 97% of males who have the fragile X have at least mild mental retardation i.e. an IQ of less than 70 (Sutherland & Hecht 1985, p.115). The vast majority of the males have moderate to severe mental retardation. The degree of mental retardation can vary widely between affected brothers (Sutherland & Hecht 1985, p.115). In cross-sectional studies, the degree of mental retardation increases with age, and it is unclear whether this reflects an organic process or the effect of

institutionalization (Sutherland & Hecht 1985, p.118). Psychometric studies of males with the fragile X syndrome indicate better verbal intelligence than performance capacity (Herbst et al. 1981; Veenema et al. 1987b; Hagerman 1989). This is in contrast with both normal males and males with other forms of non-specific X-linked mental retardation who have better performance than verbal capacities (Lehrke 1972a). Affected males frequently demonstrate autistic features, and between 5 and 16% of autistic males have the fragile X syndrome (Chudley & Hagerman 1987).

Between one third and a half of all women heterozygous for the fragile X mutation are mentally retarded (Turner et al. 1980a; Chudley et al. 1983; Fishburn et al. 1983; Fryns 1986). The majority of these affected women have mild to moderate mental retardation, and there is some dispute whether they have specific cognitive deficits or not (Veenema et al. 1987b; Hagerman 1989). Intellectually normal obligate carriers may have subtle abnormalities on psychometric testing (Loesch et al. 1987). It is not known why there is such a high proportion of symptomatic heterozygotes in the fragile X syndrome. Symptomatic heterozygotes are uncommon in other X-linked recessive disorders. The cases that have been reported can usually be attributed to non-random X inactivation (Boyd et al. 1986; Nisen et al. 1986). No evidence has been reported of non-random X inactivation in the fragile X syndrome.

The fragile X syndrome was initially defined as a form of non-specific X-linked mental retardation. Subsequent studies of fragile X males have demonstrated a number of minor physical abnormalities. None of these abnormalities are specific for the fragile X syndrome: some males with non-specific X-linked mental retardation may have these secondary features but lack the fragile X (Herbst et al. 1981; Fishburn et al. 1983), and males with the fragile X syndrome may not have any of these phenotypic abnormalities, particularly in childhood (Jacobs et al. 1980; Nielsen 1983; Sutherland & Hecht 1985, p.133). Fragile X males have an increased mean birth weight (Turner et al. 1980b), but their post-natal growth is less than normal and their adult height is generally below the 50th centile (Sutherland & Hecht 1985, p.134). Young boys with the fragile X syndrome tend to look normal but in later years they may develop a prominent forehead, long nose, prominent chin, and long protruding ears (Turner & Jacobs 1984). Macro-orchidism is an almost universal feature of post-pubertal fragile X males (Turner & Jacobs 1984). Certain dermatoglyphic abnormalities have been noted in the fragile X syndrome but they are not specific or sensitive enough to be used as a diagnostic tool (Simpson et al. 1984; Loesch 1986). Some fragile X males have features suggestive of connective tissue dysplasia, such as hyper-extensible joints, mitral valve prolapse, or pectus excavatum (Hagerman et al. 1984; Chudley & Hagerman 1987).

Mentally retarded heterozygotes may display some of these clinical features but, particularly in childhood, are often of normal appearance (Turner et al. 1980a; Turner & Jacobs 1984; Fryns 1986; Loesch & Hay 1988). Intellectually normal heterozygotes may also have subtle dysmorphic features (Loesch et al. 1987).

There have been few neuropathological studies of affected males. Fragile X males have mild generalized ventricular dilatation (Veenema et al. 1987a). There has been one preliminary report of an ultrastructural study of the brain from a fragile X male which documented reduced dendritic arborization and diminished synaptic contact[2].

There are no genetic conditions specifically associated with the fragile X syndrome. There have been case reports of individuals with both the fragile X syndrome and Klinefelter syndrome (Fryns et al. 1984a), triple-X syndrome (Fuster et al. 1988), and neurofibromatosis (Mitchell et al. 1985). The fragile X syndrome is a common disorder and is hardly surprising that an occasional individual will have both the fragile X syndrome and some other disorder. There has been only one report of the fragile X syndrome occurring in conjunction with a second X-linked disorder. Suthers et al.

---

2 Presented by K. Wyzniewski at the Fourth International Workshop on the X-linked Mental Retardation and the Fragile X held in New York, July 1989.

(1988 [Appendix D]) described a man with the fragile X syndrome and X-linked nephrogenic diabetes insipidus. Although the loci for the fragile X syndrome and nephrogenic diabetes insipidus are located close to each other on the X chromosome, DNA studies indicated that this concurrence was a chance event and was not due to a single mutation.

THE FRAGILE X SYNDROME - Cytogenetics

Fragile sites are non-staining gaps on chromosomes which can be induced under specific culture conditions. The sites are consistent in successive cultures from one person, and are inherited in a Mendelian co-dominant fashion. Fragile sites occur with differing frequencies throughout the genome, and have been classified on the basis of the manipulations of culture media required to induce them (Sutherland & Hecht 1985). Some fragile sites are expressed in less than 1% of the population, and are classified as rare. The fragile X is a rare folate-sensitive fragile site. Other fragile sites are common and are expressed by the majority of the population. Common fragile sites are induced weakly by the same folate-deficient conditions that induce expression of the fragile X, but are more strongly induced by other manipulations such as the addition of aphidicolin (Glover et al. 1984).

The fragile X is the only fragile site known to be associated with a particular disorder. It is located at Xq27.3. Under the scanning electron microscope there is a loss of the organized structure of the chromosome at the fragile site. The site is a gap of approximately 100nm and is traversed by 25-30nm chromatin strands (Harrison et al. 1983).

Although folate-deficient media were first used to induce the fragile X, a variety of special media and anti-metabolites are also effective (Sutherland & Hecht 1985, p.59). These culture manipulations all provide a stressful environment for cells which is reflected in the relatively poor cell growth. The end result of all these culture manipulations is probably a reduction in the concentration of either of the pyrimidines i.e. dTTP or dCTP (Kunz 1982; Sutherland & Hecht 1985, p.80; Sutherland & Baker 1986a; Sutherland 1988). The fragile X can be induced in lymphocytes, fibroblasts (Steinbach et al. 1983a, 1983b; Sutherland & Baker 1986b), or somatic cell hybrid lines containing a fragile X chromosome in a rodent background (Nussbaum et al. 1983; Ledbetter et al. 1986b).

The fragile X may appear in up to 60% of cultured lymphocytes from affected males. The proportion changes on repeated sampling and under different culture conditions, indicating that some of the variation in expression is related to non-genetic factors (Sutherland & Hecht 1985, p.10; Steinbach et al. 1983a). On the other hand there is a correlation in

the degree of expression between affected brothers suggesting that genetic factors also play a part (Soudek et al. 1984; Hecht et al. 1986). It is not known why this variability in expression frequency occurs, or in fact why the fragile X is not noted in every cell of an affected male. The frequency of expression of the fragile X does not correlate with the degree of mental retardation (Turner & Jacobs 1984) or other physical features in affected males (Soudek et al. 1984).

Approximately 50% of obligate heterozygotes express the fragile X (Sutherland & Hecht 1985, p.54; Fryns 1986). Heterozygotes who are mentally retarded are much more likely to express the fragile X than their intellectually normal counterparts, but the distinction is not absolute (Turner & Jacobs 1984; Fryns 1986). There is no correlation between the frequency of fragile X expression and the degree of mental retardation among fragile X positive women (Turner & Jacobs 1984). Some authors have suggested that the chromosome with the fragile X is more likely to be active in mentally retarded carriers than in intellectually normal carriers (Uchida & Joyce 1982; Uchida et al. 1983; Knoll et al. 1984) but the literature is conflicting (Fryns et al. 1984b). From a technical point of view, it is difficult to correlate X inactivation with expression of the fragile X because agents which are used to determine the replication status of the X chromosome also inhibit expression of the fragile X (Sutherland & Hecht 1985, p.122). Furthermore, the pattern of

X inactivation noted in hemopoietic cells may not correlate with the pattern in non-hemopoietic tissues such as fibroblasts. A recent study of 12 heterozygotes suggested that mental retardation was associated with an active fragile X chromosome in fibroblasts but not in hemopoietic cells. If the pattern of X inactivation in fibroblasts reflects that in neural tissue, the proportion of active fragile X chromosomes in neural tissue may dictate the degree of mental retardation in heterozygotes (Rocchi et al. 1990).

Expression of the fragile X in males is not always associated with mental retardation. Instances have been reported of intellectually normal males who are fragile X positive and who have mentally retarded fragile X positive male relatives (Webb et al. 1986b; Veenema et al. 1987a; Mulley et al. 1988). Voelckel et al. (1989) have reported a pedigree with fragile X positive men and women and no mental retardation in either sex.

The puzzling observation of mentally normal males with the fragile X has possibly been explained by the observation that there is a common fragile site close to the fragile X (Ledbetter et al. 1986a; Ledbetter & Ledbetter 1988). This common fragile site lies just proximal to the fragile X, and can be induced in both normal males and fragile X males (Ledbetter & Ledbetter 1988; Sutherland & Baker 1990). The common fragile site can be distinguished from the fragile X

on the basis of careful cytogenetic examination of banded chromosomes (Sutherland & Baker 1990), the different media conditions which best induce expression of the two sites (Ledbetter & Ledbetter 1988), and by observing the different distribution of DNA probes derived from Xq27 around both fragile sites (Sutherland & Baker 1990). Some intellectually normal men supposedly with the fragile X may have been misclassified due to induction of the common fragile site. However, recognition of the common fragile site does not resolve all the puzzling aspects of fragile X expression. On inspecting the karyotypes from the pedigree reported by Voelckel et al. (1989), the mentally normal men and women express the fragile X, not the common fragile site. Karyotypes often are not presented in the literature and it is not possible to determine how frequently the two fragile sites have been confused.

The delineation of the cytogenetics of the fragile X syndrome had major implications for the genetic counselling of families with mentally retarded children (Turner et al. 1986). It was now possible to make a specific diagnosis of an X-linked condition and to detect half of the carriers. Once the fragile X could be induced in fibroblasts it was possible to offer prenatal diagnosis (Webb et al. 1987). However, many of the possible carriers have very low levels of fragile X expression (Turner & Jacobs 1984). Do they have the fragile X, or are they expressing the common fragile site, or do they

have non-specific chromosome breakage (Jenkins et al. 1986; Mariani 1989)? The inclusion of cytogenetic data in the estimation of carrier risk estimates has reduced some of the uncertainty, in the sense that the range of carrier risks given to a woman may be narrowed. But cytogenetic identification of people heterozygous or hemizygous for the fragile X mutation can be difficult.

THE FRAGILE X SYNDROME - Population and Segregation Studies

There have been a number of prevalence surveys of the fragile X syndrome (Turner et al. 1980a; Blomquist et al. 1982; Webb et al. 1986a; Turner et al. 1986; reviewed by Webb 1989). These surveys were based on total ascertainment of children with various degrees of mental retardation in special schools. Estimates of the prevalence of fragile X males ranged from 0.4/1000 to 0.9/1000 in predominantly Caucasian populations. The prevalence of mentally retarded fragile X females was 0.2/1000 to 0.6/1000. On the basis of these figures, the fragile X syndrome is the most common cause of familial mental retardation, and is second only to Down syndrome (trisomy 21) as a cause of all mental retardation. The fragile X syndrome is not limited to just Caucasian populations, and has been documented in a wide variety of ethnic and racial groups (Webb 1989).

Segregation studies of the fragile X syndrome are difficult for a number of reasons. Mild mental retardation is not a dramatic phenotypic feature and may be missed during family studies. The variable expression of the fragile X makes it difficult to positively identify carriers, and women of unknown carrier status who express the common fragile site may be misclassified. Because the syndrome lacks characteristic physical features, a retarded female who is fragile X negative may be misclassified as having the fragile X syndrome while in fact she is a phenocopy.

Three large segregation studies of the fragile X syndrome have been carried out by Sherman (Sherman et al. 1984, 1985, 1988b). The underlying assumptions of these analyses were that the fragile X syndrome is a Mendelian disorder and that adequate correction was made for ascertainment bias. The estimation of parameters such as mutation frequencies in ova and sperm, gene frequencies, and penetrance in males and females are interdependent and rely on accurate estimates of prevalence and reproductive fitness, and accurate identification of affected individuals.

The conclusions of these studies were striking and have yet to be explained. The penetrance of the fragile X mutation (defined as the presence of mental retardation) varied according to the sex and phenotype of the parent transmitting the mutation. The penetrance among sons and daughters of phenotypically normal carrier females was 0.76 and 0.32

respectively. This indicates that phenotypically normal males can carry the fragile X mutation. Pedigree studies, including the original report by Martin and Bell (1943), confirmed that a male could be the only common relative of mentally retarded individuals who expressed the fragile X. These non-penetrant males will be referred to as "transmitting males". The penetrance among the brothers and sisters of transmitting males was 0.18 and 0.10, much lower than expected. Among daughters of transmitting males the penetrance was less than 0.01 (Sherman et al. 1985.; Froster-Iskenius et al. 1986) indicating that the sex of the carrier parent influenced penetrance. If the mutation was transmitted by a retarded carrier mother, the penetrance figures for sons and daughters were 1.0 and 0.55 respectively.

If the parental phenotype was ignored, the penetrance figures for males and females were 0.79 and 0.37 respectively (Sherman et al. 1985). On the basis of these values and taking the prevalence of the fragile X syndrome in males to be 0.44/1000, Sherman estimated the fragile X allele frequency to be 5.5×10^{-4}. This is consistent with the observed prevalence of affected females.

Estimation of the mutation rate relies on the fertility of affected males and females as measured by family size. Fragile X males are effectively infertile. Mentally normal heterozygotes have normal fertility and retarded females as a

group have reduced fertility. However carriers with borderline mental retardation have increased fertility (Sherman et al. 1984; Loesch & Hay 1988). Sherman (Sherman et al. 1984) regarded the retarded carriers as a single group with reduced fertility and estimated that the mutation rate was 7.2×10^{-4}. This is an order of magnitude higher than the mutation rates in other common X-linked conditions. This mutation rate is equal to the pooled mutation rate for all causes of non-specific X-linked mental retardation (Morton et al. 1977; Herbst & Miller 1980) and would suggest that there are fewer brain-specific loci on the X chromosome than previously suggested.

The high mutation rate proposed by Sherman et al. has been challenged (Vogel 1984; Vogel et al. 1985). Vogel has argued that the high prevalence of the fragile X syndrome may be explained by larger family sizes of heterozygotes rather than a high mutation rate. Women who are mildly retarded for any reason may have large families for social reasons (Penrose 1938, p.12; Vogel 1984). However there may also be a biological basis for increased family size among fragile X carriers. There is some evidence that the incidence of twinning is increased among carriers (Fryns 1986; Sherman et al. 1988a). The issue of what is the true mutation rate in the fragile X syndrome is unresolved.

In the initial segregation studies (Sherman et al. 1984, 1985) there was no evidence of sporadic males (i.e. fragile X males without affected relatives), implying that all the mothers of fragile X males were carriers. The corollary of this was that all fragile X mutations occur in sperm. A later study (Sherman et al. 1988b) found that 17% of fragile X males were sporadic. Thus some fragile X males are the result of new mutations that occurred during oogenesis. At this stage it seems that the mutation rate may be higher in sperm than in ova, but there is insufficient data to be confident of this. The distribution of mutations between sperm and ova does not alter the overall mutation rate.

The fragile X is one of many rare folate-sensitive fragile sites on human chromosomes (Sutherland & Hecht 1985, p.5). Unusual segregation ratios have been noted for the rare autosomal folate-sensitive sites as well as for the fragile X. In a pooled analysis of 11 rare autosomal fragile sites the penetrance of fragile site expression varied according to the sex of the transmitting parent (Sherman & Sutherland 1986). If the transmitting parent was male, the penetrance was 0.50; if the transmitting parent was female, the penetrance was 1.0. None of these rare autosomal folate-sensitive fragile sites has been associated with a human disorder, and none has been studied as intensively as the fragile X. The observation that disturbed segregation ratios

are a feature of many rare folate-sensitive sites suggests that they may have a common molecular mechanism.

* * * * *

Trying to rigorously define and analyze the fragile X syndrome was like trying to catch a shadow. Some things were clear. It was clear that mutations at one or more loci on the X chromosome could cause mental retardation without causing major congenital malformations. It was also clear that a significant proportion of males with non-specific X-linked mental retardation expressed a fragile site at Xq27.3. The consistent co-segregation of mental retardation and the fragile X in males suggested that the syndrome was due to a single locus near the fragile X, but this was not proven. It was not known whether non-specific X-linked mental retardation without the fragile X was due to allelic variation at the fragile X locus, or whether a different locus was involved. Genetic counselling had to be based on cytogenetic analysis that was at times equivocal or gave false-negative results. There was an urgent need for a different approach to research on the syndrome.

The human gene map provided a new conceptual framework for analyzing the fragile X syndrome. The development of this map is reviewed below. The techniques of molecular biology and of genetic linkage analysis furnished the necessary tools for

mapping loci of unknown function in the human genome. By localizing the fragile X locus genetically, it was possible to test whether the fragile X syndrome was genetically homogenous, to define the genotypes of individuals within fragile X pedigrees, and to directly determine the penetrance of the condition. The ultimate aim of these gene mapping studies is to isolate the fragile X mutation and to determine the molecular pathogenesis of the condition.

THE HUMAN GENE MAP - Introduction

At the turn of the century Mendelian characters were assumed to segregate independently within a pedigree, that is, the inheritance of a specific allele at one locus did not alter the probability of inheriting a particular allele at another locus. In 1905, Bateson et al. (1905) observed that two Mendelian characters in the sweet-pea did not segregate independently. The data were compatible with the two loci being linked with a recombination fraction of 0.12. At the time the authors thought these observations were at odds with the chromosomal theory of heredity. Six years later Morgan interpreted non-independent segregation and recombination in terms of exchanges between homologous chromosomes (Morgan 1911). An undergraduate student of Morgan's, A.H. Sturtevant, took Morgan's concept and examined the recombination fractions between six sex-linked characters in Drosophila

melanogaster. Sturtevant was able to determine the order and relative positions of the six loci along the chromosome. His paper (Sturtevant 1913) has become a classic. It was the first linkage map; it was a multilocus map; and it was strikingly accurate (Crowe 1988).

The first measurement of linkage in man was reported by Bell and Haldane in 1937. They estimated the frequency of recombination between the X-linked loci for color blindness and hemophilia to be 0.05 (Bell & Haldane 1937). It was Haldane who first saw the potential of gene mapping in the investigation of human health and disease. In the Croonian Lecture delivered to the Royal Society in 1946, Haldane prefaced his discussion of "The Formal Genetics of Man" with the following words (Haldane 1948):

> *"the final aim [of the formal genetic study of man] ... should be the enumeration and location of all the genes found in normal human beings, the function of each being deduced from the variations occurring when the said gene is altered by mutation ... The end result of such a genetical study as I have adumbrated would be an anatomy and physiology of the human nucleus, which would be incomparably more detailed than the anatomy and physiology of the whole body as known at present."*

To achieve Haldane's aim of localizing all human genes, two requirements had to be fulfilled. The first was the identification of many polymorphic loci in the genome that could be mapped in relation to each other. The second requirement was for a statistical methodology to efficiently analyze the linkage relationships of these loci.

THE HUMAN GENE MAP - Polymorphic loci

Up until the 1970's, the study of the linkage relationships of genetic loci was limited to those loci which had an observable phenotype. As less than 1% of the human genome consists of coding sequences (Alberts et al. 1983, p.406) a human gene map limited to the analysis of these coding regions would be a frail scaffold for Haldane's vision. However progress in molecular biology in the 1960's and 1970's yielded a technology that removed this crucial limitation to linkage analysis. It became possible to define an almost unlimited number of polymorphic loci in both the vast non-coding regions as well as the coding regions of the human genome (White & Lalouel 1986). These polymorphisms are simply variations in the DNA sequence at specific points along the chromosome. Averaged over all the autosomes, the probability that an individual is heterozygous at a specific base-pair is 0.003, i.e. a DNA sequence polymorphism occurs every 300 or so base-pairs (Cooper & Schmidtke 1984). With the development of three methodologies in molecular biology, it was possible to identify many of these polymorphisms. The first methodology was the use of enzymes which cut DNA at specific locations (restriction endonucleases)(Nathans & Smith 1975) or joining DNA molecules to permit molecular cloning (Cohen et al. 1973). The second technique was the development of laboratory methods that control the

denaturation and re-association of DNA molecules (hybridization reactions) (Marmur et al. 1963). The third critical step was the ability to rapidly incorporate radio-labelled nucleotides into short stretches of DNA (radiolabelling) (Feinberg & Vogelstein 1983).

The first polymorphisms from non-coding DNA that were used for genetic linkage studies were restriction fragment length polymorphisms (RFLPs) (Wyman & White 1980). RFLPs are due either to variations in the DNA sequence at specific restriction endonuclease sites or to the insertion or deletion of blocks of DNA between two specified endonuclease sites. The value of an RFLP for linkage studies is directly related to the degree of polymorphic variation observed at the locus. If an X-linked polymorphic locus has just two alleles, the maximum proportion of females who will be heterozygous at that locus is 0.5 (i.e. 2 x 0.5 x 0.5), and it will be possible to detect recombination around that locus in only 50% of all female meioses. If more alleles could be detected at this locus, recombination could be scored in more than 50% of female meioses in which it occurred, and the locus would be more informative. The recognition of other types of highly variable DNA polymorphisms (Nakamura et al. 1987; Noll & Collins 1987; Nobile & Romeo 1988; Litt & Luty 1989; Weber & May 1989; Vergnaud 1989; Orita et al. 1989) will make many loci informative and provide the basis for a

very high resolution genetic linkage map covering much of the human genome.

THE HUMAN GENE MAP - Two-point Linkage Analysis

The study of genetic linkage in man is plagued by problems of arbitrary pedigree structure, unselected matings, incomplete data, and - particularly in the latter part of this century - small sibships. In the 1930's and 1940's a number of statistical approaches were suggested to maximize the linkage information that could be extracted from human pedigree data (Conneally & Rivas 1980; Ott 1985, p.24; Smith 1986). The first likelihood approach to linkage analysis was made by Fisher in 1935, and was limited to two-generation pedigrees. In 1947 Haldane and Smith showed how to calculate the likelihoods of multi-generation pedigrees, but their method was particularly laborious. Smith subsequently introduced the term LODS for the logarithm of the relative likelihood[3].

Morton (1955) made the key step and introduced sequential test procedures to linkage analysis. Sequential test

---

[3] The LOD score at a particular recombination fraction is $\log_{10}$ of *the likelihood of the observed genotypes in the pedigree assuming that the loci are linked at the specified recombination fraction* minus $\log_{10}$ of *the likelihood assuming that the loci are unlinked i.e. have a recombination fraction of 0.5.* The peak LOD score indicates the best estimate of the recombination fraction in the pedigrees studied.

procedures may be used for testing hypotheses when the sample size is not fixed in advance; data are collected sequentially until there is sufficient information to accept or reject the hypothesis (Mood et al. 1974, p.464). In the case of two-point linkage analysis, Morton showed how data from separate families could be sequentially included in the analysis until there was sufficient information to either confirm linkage or to exclude it. He recommended that researchers tabulate the results of linkage analyses as LOD scores for specified values of the recombination fraction between 0 and 0.5. This method of condensing pedigree linkage data was very effective and remains in general use. The end-point for this sampling was defined in terms of the risk of falsely detecting linkage (Type I error) and the risk of falsely not detecting linkage (Type II error)(Morton 1955; Conneally & Rivas 1980). A peak LOD score of more than 3.0 (autosomal loci) or 2.0 (X-linked loci) is proof of linkage; a LOD score of less than -2 excludes linkage at that recombination fraction (Morton 1955; Conneally & Rivas 1980; Ott 1985, p.69)[4]. Morton also streamlined the application of an exact maximum likelihood method, and provided tables for the ready calculation of LOD scores for many types of two-generation pedigrees.

---

[4] These figures are derived by Bayesian analysis of the prior probability of linkage (0.022), a prior Type I error of 0.001, and a prior Type II error of 0.05. If the LOD score is 3.0, the posterior probability of falsely detecting linkage (the Type I error) is 0.05; if the LOD score is -2 the posterior probability of falsely excluding linkage (Type II error) is <0.01 (Morton 1955; Conneally & Rivas 1980).

In 1971 Elston and Stewart described a general method for computing the likelihood of a pedigree of arbitrary size and structure. The implementation of Morton's concepts and the algorithm of Elston and Stewart in a computer program, LIPED (Ott 1974; Ott 1985, p.86), made it feasible to do two-point linkage studies in any number of families of arbitrary size and structure. LIPED made allowance for qualitative or quantitative phenotypes, incomplete penetrance, age-of-onset corrections, mutation, allelic association, and missing data. LIPED was adapted to run on small personal computers, making linkage of complex pedigrees feasible for many research workers around the world.

It is possible to provide an estimate of the reliability of a recombination fraction estimated by two-point linkage analysis. For certain defined pedigree structures there are analytical methods for determining the standard error of the recombination fraction (Ott 1985, p.41). This is usually not applicable as most pedigrees have an arbitrary structure. If there is definite evidence of linkage, an approximate 90% confidence interval for the recombination fraction is given by the range of recombination fractions with LOD scores within 1.0 of the maximum LOD score (Conneally et al. 1985).

Incomplete penetrance and misclassification[5] reduce the efficiency of the linkage analysis, particularly at low recombination fractions (Ott 1985, p.124,132). The efficiency is a statistical measure of how much linkage information can be derived from the study. In the presence of incomplete penetrance, more pedigrees must be studied to confirm or exclude linkage than if the penetrance was complete. In general, mutation has little effect on the efficiency of an analysis. If the true recombination rate is zero or greater than 0.001, the loss of information due to a mutation rate of 10^{-4} is less than 5% (Ott 1985, p.127 Eqn 6.10). However, if the true recombination rate is of the same order as the mutation rate, the loss of information becomes significant because it is impossible to differentiate between a very rare recombination event and a new mutation. The sample size in human genetic linkage studies is usually inadequate to document a recombination fraction in the range 0.000-0.001,

---

5 In considering the impact of reduced penetrance on linkage analysis, it does not matter whether the incomplete penetrance is due to genetic factors or is in fact due to non-genetic factors (Greenberg & Hodge 1989). An individual may be misclassified on the basis of an incorrect diagnosis, being a phenocopy of the condition under investigation (equivalent to unrecognized linkage heterogeneity), or having a new mutation.

and in practical terms mutation has little effect on the information that can be derived from a linkage study[6].

The accuracy of the results of linkage analysis obviously depends on the accuracy of the specified parameters (and the pedigree/genotype data). If either the penetrance or allele frequencies are incorrectly specified, the recombination fraction may be biased with little change in the peak LOD score (Clerget-Darpoux et al. 1986). As a corollary, the peak LOD score may give no indication whether appropriate parameters were chosen for the linkage study. [For a contrary view, see Greenberg 1989 with criticism by Elston 1989.]

In man, as in many species, there are sex specific differences in the estimated recombination fractions. Over the whole genome recombination is more common in ova than in sperm, although in some regions this ratio is reversed (Ott 1985, p.97; White & Lalouel 1986). LIPED is able to calculate sex specific LOD scores. Obviously this consideration does not apply in linkage analysis of X-linked loci.

By using LIPED to analyze the linkage of a disease gene to another locus in a series of pedigrees, it may be possible to

---

6 It is worth noting that it is now feasible to genotype individual sperm at specified loci (Li et al. 1988). Once this process is automated, it will be possible to do linkage studies with data from millions of meioses and to document very low recombination fractions. The possible loss of information due to mutation will be offset by the vast amount of data available.

document interfamilial variability due to genetic heterogeneity. If a phenotypically homogenous condition is unequivocally mapped to different genetic locations in different pedigrees, the condition must be genetically heterogenous. Often the linkage data are not unequivocal. Ott has detailed a procedure, the A-test, for estimating the probability that genetic heterogeneity exists on the basis of two-point linkage data from a series of pedigrees (Ott 1985, p.105). The A-test examines for the presence of two types of pedigree, one type demonstrating linkage and one failing to demonstrate linkage. It is properly used to determine whether genetic homogeneity can be excluded where there is no *a priori* reason to divide the pedigrees into two groups. For this reason the test is referred to as a homogeneity test rather than a heterogeneity test. The A-test has been implemented in the program HOMOG[7]. A different homogeneity test, the B-test, has been described that has slightly greater power (in the statistical sense)(Risch 1988). It is used in the same situation as the A-test. In a situation where the pedigrees can be divided into separate groups prior to linkage analysis, Morton has described a homogeneity test for deciding whether the different groups are homogenous in terms of the recombination fraction between two loci (Morton 1956).

---

[7] The precise format of the A-test depends on the alternative hypothesis being tested (as will be detailed in Chapter 2). Ott has presented a series of HOMOG programs that can be used to test a variety of hypotheses.

These three homogeneity tests are inherently conservative, i.e. the probability of a Type I error (falsely detecting heterogeneity) is in fact less than the specified p-value. To circumvent this problem, Risch has tabulated critical values of the test statistics for each of the tests under various conditions (Risch 1988). However Risch's tables are not generally applicable, and apply only under the specified conditions.

Even if allowance is made for the conservative nature of these tests, they all require a large amount of data to detect or exclude linkage heterogeneity (Cavalli-Sforza & King 1986; Ott 1986; Risch 1988). The power of a homogeneity test varies according to the alternative hypothesis being tested. For example, a sample of 38 two-generation families each with four children will detect linkage heterogeneity at odds of 10:1 if the alternative hypothesis is that 90% of the families demonstrate tight linkage and 10% are unlinked. However, more than 10^4 families would be required to detect heterogeneity at odds of 10:1 if the alternative hypothesis was that 90% of the families were linked at a recombination fraction of 0.20 and 10% were unlinked (Cavalli-Sforza & King 1986).

In view of the arbitrary structure of human pedigrees there is no analytical method for determining the power of a

proposed two-point linkage study. In other words, there is no simple answer to the question: Is this sample of pedigrees sufficient to demonstrate or exclude linkage? In the absence of mutation or incomplete penetrance, it is possible to estimate the number of phase-known or phase-unknown meioses that must be scored to obtain a significant LOD score for a pair of loci with a specified recombination fraction (Morton 1955). However, calculating the power of a given study to detect linkage is complex, and involves consideration of the allele frequencies at both loci, the recombination fraction between them, and the penetrance of the disease gene. A computer simulation method has been described that will estimate the power of a proposed linkage study under the assumption of complete penetrance (Boehnke 1986). Although the reduction in efficiency of a two-point analysis due to incomplete penetrance (Ott 1985, p.134) or misclassified individuals (ibid. p.124) can be calculated, it is only recently that estimates of the power of a linkage study under these conditions have been calculated (Martinez & Goldin 1989). If the disease locus is identified as a quantitative trait, it is possible to estimate the power of a linkage study in relation to the measured difference in the trait in normal vs. affected family members (Demenais et al. 1988).

THE HUMAN GENE MAP - Multipoint Linkage Analysis

With the increasing number of polymorphic loci available for linkage analysis there was great interest in analyzing the linkage relationships of three or more loci simultaneously. Multipoint linkage analysis retains haplotype information i.e knowledge about the phase of three or more loci; this information would be lost in a series of two-point linkage analyses. Whether or not phase can be inferred, multipoint linkage analysis is statistically more efficient at using the information about all loci in the analysis (Lathrop et al. 1984, 1985). It provides a more accurate and precise genetic map with greater evidence for or against a particular locus order or locus position. It is a particularly valuable approach in mapping rare recessive or dominant conditions, and is much more efficient in analyzing disorders with reduced penetrance (Lathrop et al. 1985). Multipoint risk analysis is also useful for estimating genetic risks when there are no polymorphic loci tightly linked to the disease locus. The statistical efficiency of multipoint analysis increases as more loci are included.

A number of multipoint linkage analysis algorithms and computer programs have been described (Lathrop et al. 1985; Clayton 1986; Buetow & Chakravarti 1987; Lange et al. 1987). The LINKAGE package of programs (Lathrop et al. 1985; Lathrop & Lalouel 1988) has been used widely because it can be

readily adapted to address a variety of problems in genetic linkage and can be implemented on a personal computer. The LINKAGE package can essentially perform four functions. Firstly, the most likely location of a locus on a pre-defined genetic map of other loci can be estimated. Secondly, the likelihood of different locus orders (or other genetic parameters) can be calculated along with the recombination fractions between adjacent loci. Thirdly, multipoint risk analyses can be performed once the location of the disease locus is defined in relation to other loci. Finally, two-point LOD scores can be calculated. However LINKAGE (as implemented on a personal computer) is unable to analyze more than 6 loci simultaneously in pedigrees of arbitrary structure because of software and operating system constraints.

There are three major issues to be addressed in considering multipoint linkage analysis. The first two issues relate to the lack of sequential test procedures in multipoint linkage analysis; the third deals with the relationship between recombination fractions and genetic distance.

In contrast to two-point linkage analysis, none of the currently available multipoint linkage programs use sequential test procedures. Thus *there is no easy means of subsequently incorporating the results of other multipoint analyses*. When mapping the location of a disease locus on a

pre-defined genetic map (the first function referred to above), it would be feasible to sequentially add location scores[8] provided other researchers used the same pre-defined genetic map and calculated location scores at the same intervals on such a map. I am not aware of any published reports where this has been done. In practice, there is always some uncertainty regarding the accuracy of the genetic map, and it is usually simpler to share the raw pedigree and genotype data. When ordering a number of loci (the second function referred to above), subsequent data may be added if the data have been recorded in summary form as two-point LOD scores. A multipoint map can be generated from these two-point linkage data (Morton & Andrews 1989). The major penalty of this approach is that most of the information for ordering loci (i.e. haplotypes) is lost (White & Lalouel 1986).

In an attempt to circumvent these difficulties an international collaborative effort to develop the human genetic linkage map has been initiated by the Centre d'Etude Polymorphisme Humain (CEPH) (White & Lalouel 1986, 1988; Dausset et al. 1990). DNA samples from 40 three-generation

---

8 A location score is analogous to a two-point LOD score. It gives a measure of the relative likelihood that the test locus is located at a specified point on the genetic map. The peak location score indicates the most likely location of the test locus. The location score may be defined as *$\log_{10}$ of the likelihood that the test locus is located at that point* minus *$\log_{10}$ of the likelihood that the test locus is not located on the pre-defined map*. The location score may also be expressed in terms of $2\ln$(likelihood).

pedigrees each with 8 or more children are distributed to collaborators on the understanding that new genotype data will be contributed to the genotype database. The database and associated software are available to collaborators. Raw pedigree and genotype data are thus shared among investigators and duplication of effort is avoided. Complete three-generation pedigrees with many children are highly efficient in providing genetic linkage data (White et al. 1985). Linkage analysis is performed with modified versions of the LINKAGE programs which exploit the fixed structure of these pedigrees and speed the calculations by a factor of 10 or more (Lathrop et al. 1986; White & Lalouel 1988; Lathrop & Lalouel 1988). Despite these modifications analysis of many loci can take an inordinate amount of computer time, even on a mainframe computer, and more efficient algorithms and implementations have been described (Lander & Green 1987; Lathrop et al. 1988; White & Lalouel 1988; Lange & Weeks 1989).

The second feature of sequential test procedures that is lacking in multipoint linkage analysis is a clear endpoint for the analysis. At what point can you be 95% sure that the most likely locus position or order of loci is correct? Multipoint linkage analysis is statistically and numerically complex and there is no generally accepted method for determining the reliability of the results (Lathrop et al.

1984)[9]. In mapping a disease locus on a pre-defined genetic map, there is no defined relationship between the location score, the prior risk of linkage, and the Type I and II errors as is the case with two-point linkage analysis. Similarly, there is no standard statistical technique for determining a confidence interval for locus position. Computer simulation methods have been suggested (Keats et al. 1989), but these lack general applicability and would be time consuming. In practice the confidence interval for locus position is often taken to be the range of genetic locations with relative $\log_{10}$ location scores within 1.0 of the maximum location score. The significance level (in terms of a p-value) of this confidence interval is unclear (Keats et al. 1989). A variation on this method has been suggested for determining the confidence intervals for a number of recombination fractions between adjacent loci in a fixed order (White & Lalouel 1988).

The lack of a method for defining a confidence interval for gene location is particularly acute in mapping loci where there are very limited data e.g. rare disease genes, or heterogenous disorders (such as non-specific X-linked mental

---

9 This dilemma is not new. In 1949 Barnard stated that *"we statisticians have a task here of educating the scientific public that the answer to a quantitative question may consist in more than one number; for three alternatives, the position is specified by two likelihood ratios instead of one ... We must push on with this, until people are ready to accept answers in the form of sets of numbers, or even whole functions."* (cited by White & Lalouel 1986).

retardation) where linkage studies have to be limited to single families. The reliability of multipoint genetic risk estimates that might be provided when counselling such a family is critically dependent on the confidence with which the gene location is known (Lange 1986). In the absence of a confidence interval for gene location, it is difficult to specify the uncertainty of risk estimates provided to consultands.

In determining the probable order of just three loci, a test for locus order has been proposed (Ott & Lathrop 1987). When testing the order of more than three loci, two methods have been suggested to indicate how confident one can be of the most likely order. One approach is to test the likelihoods of a series of orders with inversion of adjacent loci (White & Lalouel 1988). While this may indicate that these inversions are unlikely, it is not possible to assign a significance level to this conclusion and ambiguity may persist. A second approach is to use the boot-strapping (or resampling) technique (Lathrop et al. 1987; Wilson & La Scala 1989). This can provide a significance level but involves a great deal of computer time.

The third issue to be addressed in multipoint linkage analysis is the choice of a mapping function. Multilocus linkage analysis involves the construction of a linear genetic map from recombination data, and requires a specified

relationship (or mapping function) between recombination fraction and genetic distance. A number of mapping functions have been described, the main difference between them being the degree to which they incorporate interference (Ott 1985, p.6). Interference is well documented in controlled matings of laboratory animals where it is possible to use vary large sample sizes. The appropriate mapping function for multipoint linkage studies remains a matter of some dispute in the literature. There are a number of arguments for not including interference in multipoint linkage studies (Lathrop et al. 1985; White & Lalouel 1986). Firstly, the bias in assuming no interference is well within the error range of recombination estimates for the relatively small samples sizes that can presently be achieved in human studies. Secondly, as more loci are mapped the distance between adjacent loci will get smaller, and the bias of assuming no interference will be reduced. Thirdly, ignoring interference greatly simplifies the development and application of multipoint analysis programs. Finally, the degree of interference can vary according to chromosome location (White & Lalouel 1986). The contrary view is that including interference provides a much more consistent genetic map, more accurate genetic distances, more efficient analysis, and greater support for correct locus orders (Morton et al. 1986; Pascoe & Morton 1987). It will be possible to test a variety of mapping functions in humans by using the CEPH database of pedigrees and genotypes.

As might be expected, multipoint linkage analysis is more efficient than two-point analysis in detecting genetic heterogeneity (Lander & Botstein 1986; Martinez & Goldin 1989). The A-test (Ott 1985) can be readily adapted for use with multipoint linkage data (as detailed in Chapter 2), and examines for homogeneity of locus position on a specified background genetic map. However, the significance of the result (in terms of a p-value) is not known, and it can only be expressed in terms of the odds in favor of one hypothesis vs another (Clayton et al. 1988; J Ott, personal communication). Clayton et al. (1988) have described a similar multipoint homogeneity test.

* * * * *

Genetic linkage analysis is not the only method for mapping loci in the human genome. The use of somatic cell hybrid lines and *in situ* hybridization are two techniques that physically position loci on chromosomes. When somatic cell hybrid lines were first used for locus mapping, a panel of cell lines containing a variety of human chromosomes or chromosome fragments was assayed for the protein of interest. By comparing which cell lines expressed the protein it was possible to localize the corresponding gene to a particular chromosomal region. By using sophisticated selection mechanisms it is now possible to generate hybrid lines that contain a single discrete fragment of a human chromosome

(Rosenstraus & Chasin 1975; D'Urso et al. 1983; Callen 1986; Brown & Willard 1988; Benham et al. 1989; Warburton et al. 1990). A panel of cell lines containing overlapping fragments allows high resolution physical mapping of DNA probes (Wieacker et al. 1984; Murphy & Ruddle 1985; Riddell et al. 1986; Oberle et al. 1986a; Patterson et al. 1987b; Cremers et al. 1988; Schonk et al. 1989).

In situ hybridization (Simmers et al. 1988) of a labelled DNA probe to banded chromosomes allows direct visualization of the probe's location. This procedure is technically demanding and provides relatively low resolution. However, it may be more rapid than an extensive linkage study or mapping with cell lines, and is particularly valuable in localizing DNA probes in relation to cytogenetic abnormalities such as translocation breakpoints or fragile sites.

The application of the advances in molecular biology and in genetic linkage methods to development of the human gene map was recognised in a seminal paper by Botstein et al. in 1980. They suggested that a genetic map of the non-coding regions of the genome could be developed with just several hundred polymorphic loci (Botstein et al. 1980; Lange & Boehnke 1982). Within 7 years sufficient RFLPs had been characterised and mapped to generate a genetic linkage map covering 95% of the human genome (Donis-Keller et al. 1987). With the inclusion of much of the non-coding region of the human

genome in the gene map, the tenuous linkage map of the 1970's has become a solid foundation for Haldane's vision. It has been possible to localize genetic loci causing diseases without any prior knowledge of the function of the responsible genes. Localization of disease genes has allowed for precise genetic counselling (Gusella 1986) and for the isolation and cloning of a number of disease genes (Lange et al. 1985; White & Lalouel 1988). Once a high resolution map of the genome is defined it will be possible to map the various loci responsible for polygenic disorders (Paterson et al. 1988) or the loci involved in complex genetic traits (White & Lalouel 1988).

THE FRAGILE X SYNDROME - Linkage Analysis

Prior to the identification of polymorphisms in non-coding regions of DNA, there were only two X-linked non-pathogenic polymorphisms that could be used for genetic linkage studies. The Xg blood group locus was located at distal Xpter while the glucose-6-phosphate dehydrogenase (G6PD) locus was located at distal Xqter (Davies 1985). Filippi et al. studied the linkage relationship of G6PD and the fragile X mutation,

FRAXA[10], in a series of Sardinian pedigrees (Filippi et al. 1983). The maximum LOD score was 3.16 at a recombination fraction of 0.06. This result was consistent with the suggestion that FRAXA was close to the fragile X (Kaiser-McCaw et al. 1980).

With the development of a genetic linkage map of the X chromosome using DNA polymorphisms (Drayna & White 1985) it was possible to explore linkage relationships of FRAXA and polymorphic loci located at Xqter. The Hemophilia B locus (F9) was initially reported to show tight linkage to FRAXA (Camerino et al. 1983). This was later disputed (Choo et al. 1984; Warren et al. 1985) and F9 was shown to be proximal to FRAXA (Davies et al. 1985). The recombination fraction between F9 and FRAXA is 0.21 (Keats et al. 1989). In 1984 the probe 4D-8 (DXS98)[11] was isolated (Boggs & Nussbaum 1984) and was subsequently shown to also lie proximal to FRAXA with a

---

10 The terms 'fragile X syndrome', 'fragile X', and 'FRAXA' are not synonymous. The symbol FRAXA refers to the abnormal DNA responsible for the fragile X syndrome. The term 'fragile X' refers to the rare folate-sensitive fragile site at Xq27.3. The 'fragile X syndrome' may be defined as the combination of mental retardation and expression of the fragile X.

11 A DNA probe hybridizes to a specific locus in the genome. Linkage analysis indicates the relationships between loci, not between probes. Probes have a variety of laboratory names (e.g. 4D-8) but loci have specific designations that are written in italics or underlined (e.g. DXS98). In practice, the two terms are often used interchangeably. In this thesis, locus symbols will generally be used for established loci or cloned genes, and probe names will generally be used for loci mapped as part of this project.

recombination fraction of 0.10 (Brown et al. 1987a; Keats et al. 1989). DXS98 is not a very informative locus, and a more informative locus proximal to FRAXA, DXS105 (cX55-7), was identified in 1987 (Hofker et al. 1987; Arveiler et al. 1988a; Rekila et al. 1988). The recombination fraction between DXS105 and FRAXA is 0.16 (Keats et al. 1989).

On the distal side of FRAXA, the probe St14-1 (DXS52) was isolated in 1985 (Oberle et al. 1985; Heilig et al. 1988) and has a recombination fraction of 0.14 (Keats et al. 1989). DXS52 is a highly polymorphic locus. Although other loci have been identified which lie between FRAXA and DXS52 (Patterson et al. 1987a; Patterson et al. 1989; Vincent et al. 1989a) they are tightly linked to DXS52 and are of little additional benefit for either developing the genetic map around FRAXA or for providing genetic counselling based on DNA studies.

In 1989 two new polymorphic loci within 5cM of FRAXA were reported. DXS369 (RN1) lies approximately 5cM proximal to FRAXA (Hupkes et al. 1989; Oostra et al. 1990) while DXS304 (U6.2) lies 3cM distal to FRAXA (Dahl et al. 1989a, 1989b; Vincent et al. 1989b; Rousseau et al. 1990).

Two-point linkage studies of fragile X pedigrees and a major multipoint linkage analysis of FRAXA (Brown et al. 1988) were summarized at the biennial Human Gene Mapping Workshop held in 1989 (Mandel et al. 1989; Keats et al. 1989). The order of

loci in the region of FRAXA is shown below. The genetic distances between adjacent loci (in cM) were derived using Haldane's mapping function[12] from the recombination fractions listed in a number of sources (Keats et al. 1989; Vincent et al. 1989a, 1989b; Oostra et al. 1990).

```
                                                      St35/
          cX55-7   4D-8    RN1              U6.2      IA1     St14-1
cen-------------------------------------------------------------------qter
~~qter~~
    F9   DXS       DXS     DXS     FRAXA    DXS       DXS     DXS
         105        98     369              304       305/     52
                                                      374
cM     8         8       6      5         3        12       1
```

The genetic distances given above are based on two-point linkage analyses in different sets of fragile X pedigrees. A multipoint linkage study of these loci in the CEPH pedigrees has yet to be reported.

FRAXA is consistently mapped to Xq27 with no evidence of the fragile X locus being unlinked to these markers in any pedigrees. There is one possible exception to the apparent genetic homogeneity of the fragile X syndrome. It has been suggested that the linkage relationship between F9 and FRAXA is not homogeneous in fragile X pedigrees, and that some families have tight linkage between F9 and FRAXA while others show little or no linkage (Brown et al. 1985, 1986, 1987b,

---

12 $x = -0.5\ln(1-2r)$ where x is the genetic distance in cM, and r is the recombination fraction (Ott 1985, p. 8). Haldane's mapping function makes no allowance for interference.

1988). This would explain the contradictory reports that were initially published describing the F9-FRAXA linkage relationship. Using both the A-test (Ott 1985, p.105) and the B-test (Risch 1988) on the one set of fragile X pedigrees, the linkage relationships between F9-FRAXA and F9-DXS52 were shown to be heterogenous (Brown et al. 1987b; Risch 1988). There was no evidence of heterogeneity between FRAXA and DXS52 suggesting that the heterogeneity reflected different recombination fractions proximal to FRAXA, and was not due to the fragile X locus being located away from Xq27 in some pedigrees. Approximately 20% of fragile X pedigrees had no recombination between F9 and FRAXA, while the remainder demonstrated a recombination fraction of approximately 0.35 (Brown et al. 1988). [In the summary map above, the genetic location of F9 in relation to FRAXA is based on pooled linkage data]. An initial suggestion that transmitting males were more common in families with tight F9-FRAXA linkage (Brown et al. 1985) has been refuted (Brown et al. 1987b).

There are a number of reasons for advocating caution in attributing this heterogeneity to the fragile X mutation. Firstly, linkage heterogeneity may be a feature of the normal gene map. Normal pedigrees have not been tested for linkage heterogeneity in this region. It has been suggested that the frequency of recombination around DXS52 may vary in normal pedigrees (Drayna & White 1985), but this has not been rigorously tested. Oberle et al. (1986b) found no

heterogeneity when comparing the F9-DXS52 linkage relationship in normal versus fragile X pedigrees. Secondly, linkage heterogeneity has not been documented for FRAXA and the proximal locus DXS98 (Brown 1989), although this probably reflects the small number of pedigrees tested (Ott 1986). Thirdly, the testing for genetic heterogeneity with the A-test or B-test involved multiple pair-wise comparisons, and it is not clear what allowance was made for this. As a conservative approach, for N pair-wise comparisons the significance level should be 0.05/N so that the overall p-value is 0.05 (J Ott, personal communication). On this basis, the degree of F9-FRAXA heterogeneity (Brown et al. 1988) is highly significant, but the degree of F9-DXS52 heterogeneity (Risch 1988) is not significant. Finally, a multipoint study of heterogeneity in fragile X pedigrees failed to confirm linkage heterogeneity (Clayton et al. 1988). F9, DXS98, and DXS52 are all more than 10cM from FRAXA (Keats et al. 1989). These studies need to be repeated in both normal and fragile X pedigrees using multipoint homogeneity tests and loci that are closer to FRAXA.

Genetic linkage studies of the fragile X syndrome have resolved some of the questions left unanswered by evaluating clinical and cytogenetic features or by segregation analysis. In the first place, the fragile X syndrome is a genetically homogenous condition i.e. it is due to a single locus, FRAXA, located at Xq27. Secondly, FRAXA is either at or very close

to the fragile X itself. Polymorphic loci which lie proximal or distal to FRAXA on linkage studies are located proximal or distal to the fragile X on *in situ* hybridization of the labelled probes to chromosomes expressing the fragile X (Szabo et al. 1984; Schnur et al. 1989; Suthers et al. 1989a [Appendix D]). Thirdly, genetic linkage studies have, to a limited extent, confirmed the penetrance figures reported by Sherman et al. (1984, 1985). In Sherman's segregation analysis the penetrance of the fragile X mutation in the sons of intellectually normal mothers was approximately 80%. This result predicts that approximately 17% of the normal sons of intellectually normal carriers would have the same haplotype for loci immediately flanking FRAXA as their mentally retarded brothers. This prediction has been confirmed (Brown et al. 1990). Finally, the development of the genetic linkage map around FRAXA has made it feasible to identify carriers and provide genetic counselling in fragile X pedigrees (Mulley et al. 1987; Sutherland & Mulley 1990).

However, as might be expected in dealing with such an unusual condition, some aspects of genetic linkage studies around FRAXA have raised further questions. First, progress in identifying polymorphic loci close to FRAXA has been very slow. Until 1989, no locus was isolated that was closer than 10 cM to FRAXA. As only one locus (DXS52) near FRAXA was highly polymorphic genetic risk estimates were often based on genotypes of loci distant from FRAXA. To some extent the lack

of closely linked polymorphic loci was due to difficulties in isolating RFLPs on the X chromosome. RFLPs are less common on the X chromosome than the autosomes (Cooper & Schmidtke 1984; Hofker et al. 1986) due to conservation of the X chromosome during evolution (Ohno 1969). As a consequence, searches for RFLPs are three times less efficient on the X chromosome than on the autosomes (Hofker et al. 1986)[13].

Second, while the fragile X syndrome is genetically homogenous, it is still not known whether other forms of non-specific X-linked mental retardation are allelic to FRAXA, or whether they are genetically distinct.

Third, the possibility of linkage heterogeneity between F9 and FRAXA in fragile X pedigrees makes it difficult to provide reliable estimates of carrier risk on the basis of F9 genotypes.

Fourth, in performing genetic linkage studies in fragile X pedigrees it is necessary to infer the genotype of an individual at FRAXA from the phenotype using parameters such as penetrance and mutation rate. The reduced penetrance of the fragile X mutation reduces the efficiency of the

---

13 A second class of DNA polymorphisms, 'variable number of tandem repeats' (VNTRs), are also relatively uncommon on the X chromosome. However, a recently-described class of highly polymorphic AC_n repeat polymorphisms is probably as common on the X chromosome as on the autosomes (Luty et al. 1990).

analysis. At a recombination fraction of 0.10, the incomplete penetrance of the fragile X mutation reduces the efficiency of a two-point linkage study by a factor of two to three (Ott 1985, p.134). Although the penetrance in sons of retarded carriers is reportedly 1.0 (Sherman et al. 1985) it is not clear what degree of retardation a carrier must have for the penetrance to be that high. If a carrier is incorrectly coded as being retarded, the penetrance in her sons will be falsely high and the recombination fraction will be over-estimated (Clerget-Darpoux et al. 1986). The uncertainty regarding the distribution of the mutations between ova and sperm makes little difference in genetic linkage studies of pedigrees with two or more affected individuals. However, this distribution is critical in providing carrier risk estimates to the mother of a sporadic fragile X male (Sherman et al. 1988b).

The fifth dilemma that genetic linkage studies of FRAXA has raised is the occasional conflict of cytogenetic and DNA studies. It is recognised that there are transmitting males who are phenotypically and cytogenetically normal despite inheriting the same haplotype of loci flanking FRAXA as their affected brothers (Brown et al. 1990). However, there have been occasional reports of mentally normal males who on linkage analysis are at low risk of having FRAXA who express the fragile X at low frequencies (Mulley et al. 1988). The explanation may be that such individuals expressed the common

fragile site rather than the fragile X (Sutherland & Baker 1990). The cytogenetic distinction between the fragile X and the common site was made only recently, and some men and women may have been misclassified as fragile X positive in earlier genetic linkage studies.

THE FRAGILE X SYNDROME - Hypotheses

The fragile X syndrome is an enigma. Numerous models have been proposed in an attempt to integrate the unusual and variable clinical, cytogenetic, and genetic features of the syndrome. The types of models that have been suggested include the following :

the fragile X syndrome is due to the interaction of several genes (Steinbach 1986; Israel 1987; Sherman 1987);

the variable fragile X phenotype is mediated mainly by the uterine environment as a consequence of different patterns of X inactivation in the uterine tissues (Van Dyke & Weiss 1986);

dermatoglyphic findings in the fragile X syndrome could indicate a submicroscopic X;Y reciprocal translocation (Loesch 1986);

the fragile X is due to disruption of normal gene function by transposons (Friedman & Howard-Peebles 1986; Hoegerman & Rary 1986);

a non-pathogenic pre-mutation at FRAXA is converted to a deleterious mutation by recombination (Pembrey et al. 1985; Winter & Pembrey 1986; Winter 1987; Brown et al. 1987c; Winter & Pembrey 1987; Schaap 1989);

the unusual cytogenetic characteristics of the fragile X could be due to repeated polypurine/polypyrimidine sequences (Sutherland et al. 1985).

These hypotheses have been reviewed by Nussbaum and Ledbetter (1986). They present their own suggestion that FRAXA is a pyrimidine-rich sequence that is amplified in a step-wise fashion in successive meioses; the amplified sequence disrupts one or more genes and causes the phenotype of mental retardation and expression of the fragile X.

Laird has proposed a sophisticated model which attributes the features of the fragile X syndrome to abnormal X inactivation and re-activation at Xq27.3 (Laird et al. 1987). This process of 'imprinting' could account for the unusual penetrance and population genetics of the syndrome (Sved & Laird 1990). On the basis of this model, Laird has also suggested that the

oogonial cells of a female are derived from just two primordial cells in the embryo (Laird et al. 1990).

These hypotheses will not be discussed further. It is possible to interpret the observed features of the syndrome in relation to a number of these models, but none of the hypotheses is readily testable. There has been one investigation of the 'imprinted X' model which found no evidence of abnormal methylation (suggesting X inactivation) at genes located at Xq27-q28 (Khalifa et al. 1990). This does not disprove the 'imprinted X' hypothesis because the genes examined were all more than 10 cM from FRAXA. Because there is no robust testable model of the fragile X mutation, there is a risk that a model could be accepted uncritically (Suthers & Sutherland 1990 [Appendix D]). Once the DNA sequence at the fragile X locus has been cloned, it will be possible to evaluate the hypotheses and to address the cell biology of this fascinating condition.

CONCLUSION

The fragile X syndrome remains a puzzle for clinicians, counsellors, and geneticists. Progress towards isolating a DNA probe specific for the fragile X mutation has been slow. This lack of progress is due in part to the reduced frequency of RFLPs on the X chromosome, and in part to the unusual

penetrance of the syndrome. The problem of reduced penetrance would be lessened if a series of highly polymorphic tightly-linked loci were identified near FRAXA.

One question that could be readily addressed with genetic linkage studies is whether other forms of non-specific X-linked mental retardation are allelic to FRAXA. 95 linkage studies indicated that such a locus was placed well away from Xq27 this would indicate that nonspecific X-linked mental retardation is genetically heterogenous. This issue is addressed in Chapter 3 where the position of the locus responsible for non-specific X-linked mental retardation in a large pedigree is described. In this family the locus (MRX1) was mapped to the short arm of the X chromosome, indicating that MRX1 is not allelic to FRAXA and that non-specific X-linked mental retardation is genetically heterogenous. In order that multipoint carrier risk estimates could be provided for the women in this pedigree, a confidence interval for MRX1 location was determined using a novel resampling strategy.

The difficulty of identifying RFLPs on the X chromosome has prompted researchers to utilize physical mapping methods. Somatic cell hybrids containing human X chromosomes with translocation or deletion breakpoints have been used to physically map loci in various regions of the long arm of the human X chromosome (Wieacker et al. 1984; Oberle et al.

1986a; Hofker et al. 1987; Cremers et al. 1988). Once a probe is shown to identify a locus in the region of interest the search for RFLPs can be focussed on that locus. However, few cell lines have been identified that have breakpoints that are close enough to the FRAXA to be useful in this regard.

A panel of cell lines with X chromosome breakpoints close to FRAXA is described in Chapter 4. This panel was used to physically localize 18 new DNA probes near FRAXA. It was possible to order both the X chromosome breakpoints and the new probes in relation to FRAXA without relying on linkage studies. Investigation of one of these cell lines indicated that the gene responsible for Hunter syndrome is located immediately distal to FRAXA.

If RFLPs can be identified at loci close to FRAXA, linkage studies can be used to confirm the order of the loci suggested by physical mapping. Such RFLPs would have an immediate role in genetic counselling. Three of the probes detected polymorphic loci very close to FRAXA (Chapter 5). Multipoint linkage studies of these new RFLPs and adjacent loci in the CEPH pedigrees confirmed the order suggested by the physical mapping studies. There was no evidence of linkage heterogeneity in these families as assessed using two-point homogeneity tests.

The multipoint linkage relationships of these new probes in relation to FRAXA were defined in a collaborative study of 112 fragile X pedigrees (Chapter 6). The three new polymorphic loci are the closest polymorphic loci to FRAXA yet described, and are the basis of an efficient strategy for diagnostic DNA studies in fragile X families. Multipoint tests of linkage heterogeneity of these loci near FRAXA did not indicate genetic heterogeneity (Chapter 7).

Chapter 2

MATERIAL AND METHODS.

INTRODUCTION

The patient DNA samples, cell lines, and DNA probes used in this project were obtained on a collaborative basis from many sources. Established methods for handling DNA were employed and are briefly described.

Linkage analyses were performed using the LINKAGE package of programs. The implementation of the LINKAGE programs is described in some detail. The HOMOG programs are described, and the data requirements and limitations of the analyses presented. The genetic parameters used in linkage analyses of the fragile X families are indicated.

A large volume of genotype data was generated during this project, and a number of small computer programs for handling these data are described. In addition programs for estimating the length of DNA fragments and for converting recombination fractions to centiMorgans are presented.

A resampling method for determining approximate confidence intervals for the results of multipoint linkage analysis is described. This method was used to define confidence intervals for the locations of MRX1 (Chapter 3) and FRAXA (Chapter 6).

DNA SAMPLES FROM PEDIGREES

Blood samples were collected from 31 members of a pedigree with non-specific X-linked mental retardation. The family was counselled prior to the collection, and individual members gave informed consent to Dr G Turner (Randwick, N.S.W.) for both sample collection and for subsequent genetic linkage analysis. DNA was extracted from peripheral lymphocytes using the phenol/chloroform method (Maniatis et al. 1982).

Phenol-chloroform extraction of DNA from lymphocytes:

Blood samples were collected into an EDTA tube. 10 ml of blood was diluted with 15 ml of cell lysis buffer (0.32 M sucrose/ 10 mM TrisHCl/ 5 mM $MgCl_2$/ 1% Triton X-100), and left on ice for 30 minutes, spun at 2000G at 4°C for 15 minutes, and the supernatant removed to the 5 ml mark; 20 ml of cell lysis buffer was added, mixed, and the solution spun at 2000G for another 15 minutes. All the supernatant was removed; 1 ml of Proteinase K buffer (10 mM TrisHCl/ 10 mM NaCl/ 10 mM disodium EDTA) was added and the solution vortexed gently; a further 2.25 ml of buffer, 0.5 ml of 10% SDS, 0.2 ml of 1% Proteinase K (Boehringer Mannheim) were added, and incubated overnight on a slow wheel at 37°C. Five ml of phenol (previously distilled, then saturated with 10 mM TrisHCl) was added, the tube was mixed gently for 15 minutes, and spun at 2000G for 10 minutes; the top aqueous phase was transferred to a 10 ml tube, 5 ml of

phenol added, the tube mixed for 10 minutes, and then spun at 400G for 10 minutes. The top aqueous phase was transferred to a fresh 10 ml tube; 5 ml of 24:1 chloroform:isoamyl alcohol was added and gently mixed before the tube was spun at 400G for 10 minutes. Then 0.3 ml of 3 M NaAcetate (pH 5.2) and cold 99% ethanol to 10 ml were added, and mixed gently until the DNA precipitated. The DNA pellet was washed well with cold 70% ethanol, desiccated, and dissolved in 0.1 ml of 10mM TrisHCl/0.1 mM EDTA. Gloves were worn throughout the extraction; phenol and chloroform were handled in a fume hood.

DNA samples from 40 normal pedigrees were provided on a collaborative basis by the Centre d'Etude Polymorphisme Humain (CEPH)(Dausset et al. 1990). DNA samples from 35 fragile X pedigrees were provided by Dr J Mulley (Adelaide, S.A.).

CELL LINES

A panel of 14 cell lines and patient DNA samples was assembled for physical mapping studies of DNA probes at Xq27-q28. The cell lines were obtained from a variety of sources on a collaborative basis. The collaboration was initiated and managed by the candidate. The cell lines are briefly

described below. The cell type, contributing co-author, and references for each cell line used in the physical mapping studies are summarized in Table 2-A. The cell lines were maintained in the laboratory by T Hocking and S Lane under the supervision of DF Callen (Adelaide, S.A.), while the candidate extracted the DNA from the cell lines.

CY2 This was a mouse-human hybrid cell line (Callen 1986). The mouse background was an HPRT$^-$ line (A9). The only human chromosome content was a derived X chromosome containing Xpter-Xq26 [karyotype der(X)t(X;16) (q26;q24)].

CY3 This was a mouse-human hybrid cell line, and was the reciprocal of CY2 (Callen 1986). It contained a derived human chromosome 16 and Xq26-Xqter [karyotype der(16)t(X;16)(q26;q24)].

CY34 This cell line was derived from a girl with Hunter syndrome [iduronate sulfatase (IDS) deficiency; mucopolysaccharidosis II] and a t(X;5)(q28;q33) reciprocal translocation (Mossman et al. 1983; Roberts et al. 1989). The translocation breakpoint was at Xq28, and the normal X chromosome had been consistently inactivated. It was postulated that IDS was located at the breakpoint on the X

Table 2-A

Cell lines used in this study.

Cell line	Cell type[a]	Source	Reference
CY2	SCH/mouse	DF Callen, Adelaide	Callen 1986
CY3	SCH/mouse	DF Callen, Adelaide	Callen 1986, Suthers et al. 1989a
CY34	SCH/mouse	DF Callen, Adelaide	Suthers et al. 1989a
PeCH-N	SCH/hamster	MC Hors-Cayla, Paris	Suthers et al. 1989a
LL556	lymphoblastoid (XY)	I Oberle, Strasbourg	Suthers et al. 1990a
TC4.8	SCH/hamster	M Rocchi, Genoa	Rocchi et al. 1989, Suthers et al. 1990a
Y.162.Aza	SCH/hamster	M Rocchi, Genoa	Rocchi et al. 1989, Suthers et al. 1990a
APC-5	SCH/hamster	N Thomas, Cardiff	Suthers et al. 1990a
04-1	lymphocytes (XY)	JJ Hopwood, Adelaide	Wilson et al. 1990
03-1	lymphocytes (XY)	JJ Hopwood, Adelaide	Wilson et al. 1990
2384-A2	SCH/hamster	CE Schwartz, S Carolina	Ledbetter et al. 1990
LC12K15	SCH/mouse	M Schmidt, Melbourne	Schmidt et al. 1990
908K1B17	SCH/hamster	HH Ropers, Nijmegen	Schonk et al. 1989, Oostra et al. 1990
GM08121	lymphoblastoid (XX)	N.I.G.M.S.[b]	Patterson et al. 1987

a SCH indicates a somatic cell hybrid line; the background cell type is shown. The sex chromosome content of lymphocytes or lymphoblastoid lines is indicated.

b U.S. National Institute of General Medical Sciences Human Genetic Mutant Cell Repository.

chromosome. The translocated X chromosome (i.e. Xpter-Xq28) was isolated in the A9 mouse cell line.

The cell line CY34A was a subclone of CY34 that had arisen spontaneously and had been detected during routine cytogenetic screening. It contained just a fragment of the original derived X chromosome extending from Xq24 to Xq28.

PeCH-N This was a hamster-human hybrid cell line. The human X chromosome content consisted of the derived X chromosome (i.e. Xpter-Xq27) from a balanced t(X;21)(q27;q11) translocation (Couturier et al. 1979).

LL556 This was a lymphoblastoid line from one of two brothers with hemophilia B (Factor 9 deficiency) and mental retardation. The boys had cytogenetically visible deletions extending from Xq26.2 to Xq27.2 (Suthers et al. 1990a [Appendix D]).

TC4.8 Peripheral lymphocytes from a male expressing the fragile site at Xq27.3 were fused (Davidson & Gerald 1976) with <u>HPRT</u>$^-$/<u>G6PD</u>$^-$ hamster cells (YH.21) (Rosenstraus & Chasin 1975). After selection for <u>HPRT</u> in HAT medium, one clone (HY.84P11) containing

a human X chromosome and no other human chromosomes was isolated. HY.84P11 was treated with FUdR and caffeine to induce expression of the fragile site (Abruzzo et al. 1986) and breakage of the X chromosome at that point. Clones retaining HPRT were selected against with 6-thioguanine; surviving clones were selected for retention of G6PD by treatment with diamide (D'Urso et al. 1983). TC4.8 was a homogeneous clone that retained G6PD (Rocchi et al. 1989). Subsequent analysis (Chapter 4) demonstrated that the breakpoint was not at the fragile X but had occurred more proximally on the X chromosome.

Y.162.Aza Y.162.Aza was derived from a somatic cell hybrid line (Y.162.SE1T4) which contained an intact late-replicating human X chromosome in a hamster background. On cytogenetic screening of Y.162.SE1T4, a proportion of clones were noted to contain an elongated human X chromosome with an additional early-replicating fragment attached at Xqter. A subclone homogeneous for this rearrangement (Y.162.Aza) was studied with *in situ* hybridization of labelled total human or total hamster DNA. The early-replicating fragment attached to the human X chromosome was shown to be

of hamster origin, and no other human chromosomal material was detected (Rocchi et al. 1989).

APC-5 A girl with mild mental retardation and no major dysmorphic features was found to have a balanced X;autosome translocation: 46,Xt(X;19)(p11.2;19q13.3). Skin fibroblasts were fused (Davidson & Gerald 1976) with an HPRT$^-$ hamster cell line (Wg3h) and clones containing the derived X chromosome were selected for HPRT in HAT medium. The human chromosomal material in the cell line was fragmented and the cell line has been only partially characterised cytogenetically (Suthers et al. 1990a [Appendix D]). Subsequent analysis (Chapter 4) demonstrated a breakpoint at Xq27 which was presumably due to fragmentation in the hybrid cell line.

03-1 and 04-1

Two boys with Hunter syndrome (iduronate-2-sulfatase [IDS] deficiency; mucopolysaccharidosis type II) due to complete deletions of IDS have been reported (Wilson et al. 1990 [Appendix D]). The boys, 03-1 and 04-1, had extremely severe features of Hunter syndrome. They had presented by one year of age with developmental delay and by the second year had developed hernias, vertebral collapse,

enlarged liver and spleen, and developmental regression. Neither boy attained speech. They did not have any congenital malformations although the severity of their Hunter syndrome phenotype made it difficult to exclude either mild congenital facial dysmorphism or congenital mental retardation. DNA was extracted from peripheral lymphocytes of these boys. For the sake of brevity in the text, the peripheral lymphocytes will be included in the term "cell lines".

2384-A2 This was a hamster-human hybrid cell line. The deleted X chromosome (i.e. Xpter-Xq27) from a woman with a deletion of her paternal X chromosome [karyotype 46,X,del(X)(q27)] was isolated in an HPRT$^-$ hamster cell line (Ledbetter et al. 1990).

LC12K15 The deleted X chromosome from a girl with an interstitial deletion at Xq27 [karyotype 46,X,del(X)(q27.1q27.3)] (Schmidt et al. 1990) was isolated in a mouse cell line.

908K1B17 This hamster-human hybrid cell line contained a single der(19;X) human chromosome. The X chromosome fragment was reported to extend from Xq24 to qter (Schonk et al. 1989). Subsequent analysis (Chapter 4) indicated that it extended from Xq27.

GM08121 This was a lymphoblastoid line from a woman with a deletion extending from Xq26 to Xqter on one chromosome (Patterson et al. 1987b).

For physical mapping of new probes with these cell lines, DNA from normal human lymphocytes, mouse (A9) cells (Callen 1986), and hamster (RJK 88) cells (Fuscoe et al. 1983) were used as positive and negative controls.

Cell line DNA was extracted using the high-salt extraction method (Miller et al. 1988).

High-salt extraction of cell line DNA. Cell pellets were frozen at -20°C, thawed, washed twice in cell lysis buffer, and incubated in buffer on ice for 30 minutes; the tube was then spun and the supernatant removed. Three ml of nuclei lysis buffer (10 mM TrisHCl/ 400 mM NaCl/ 2 mM EDTA.pH8) were added and the pellet loosened; 0.5 ml 10% SDS and 0.2 ml 1% Proteinase K (Boehringer Mannheim) were added, the tube mixed gently, and incubated at 37°C overnight. One ml of saturated NaCl (6M) was added and the tube shaken vigorously for 15 seconds; the tube was spun at 1000G at 4°C for 15 minutes, and the supernatant transferred to a 50 ml tube. Two volumes (10 ml) of 100% ethanol at room temperature were added, and gently mixed until the DNA precipitated; the DNA was transferred to a 10 ml tube

with 10 ml 70% ethanol at room temp; after gentle mixing the tube was spun and the pellet recovered; the pellet was desiccated for 5 minutes and resuspended in 2 ml of nuclei lysis buffer at 4°C overnight.

The DNA pellet was treated with RNase A (Boehringer Mannheim; 0.15 ug/ml at 37°C for 4 hours) and a further Proteinase K digestion (0.07 ug/ml at 37°C overnight). The DNA was precipitated by the addition of 1/3 the volume of saturated NaCl and 1/5 the volume of 50% PEG 6000, gently mixed, and allowed to stand at 4°C overnight; the DNA pellet was recovered by spinning at 1000G at 4°C for 15 minutes, washed with 70% ethanol, desiccated, and resuspended in 10mM TrisHCl/1mM EDTA.

DNA PROBES

A series of new DNA probes localized to Xq26->qter were provided by Dr VJ Hyland (Adelaide, S.A.). The isolation of this series of 17 DNA probes (all with the prefix "VK") has been described (Hyland et al. 1989). In brief, a genomic library was prepared from the somatic cell hybrid CY3 which contained Xq26->qter as the only human X chromosome component. The genomic library was packaged into the SalI site of the lambda vector EMBL3. Clones containing human DNA were identified by probing the library with total human DNA;

human-derived clones were mapped to either Xq26-qter or to chromosome 16 (the other human chromosome in CY3) by probing a panel of somatic cell hybrids. The insert size of the VK clones ranged from 12.2 to 23.8 kilobases (kb). The loci identified by the VK probes are listed in Table 2-B.

Table 2-C lists the probe name, locus, source, and reference for the other DNA probes used in this project. Most of these probes were obtained by Dr JC Mulley (Adelaide, S.A.).

SOUTHERN BLOTTING & RADIO-LABELLING

DNA samples were digested with one of a variety of restriction endonucleases (New England Biolabs). 10ug of DNA was digested under the conditions specified by the manufacturer with a fourfold excess of endonuclease for 16 hours[1]. The digestion was stopped by the addition of EDTA to a concentration of 10 mM.

DNA samples were electrophoresed along with standard DNA size markers in 0.8% agarose gel. The gel was then soaked in an ethidium bromide solution (0.0025 mg/ml), rinsed in water, and photographed under UV light to document the positions of

---

[1] If the restriction endonuclease buffer had a salt concentration in the range 50-100 mM, spermidine was added to a final concentration of 5 mM. If the buffer salt concentration was greater than 100 mM, spermidine was added to a final concentration of 10 mM.

Table 2-B

Loci detected by the VK series of probes

Probe name	Locus
VK7	DXS288
VK9	DXS289
VK10	DXS290
VK11	DXS291
VK14	DXS292
VK16	DXS293
VK17	DXS294
VK18	DXS295
VK21	DXS296
VK23	DXS297
VK24	DXS298
VK25	DXS299
VK29	DXS300
VK34	DXS301
VK37	DXS302
VK41	DXS310
VK47	DXS288

Table 2-C.
Probe name, locus name, location, source, and reference for probes used in the study.

Probe	Locus	Location	Source	Reference
RC8	DXS9	Xp22	KE Davies, Oxford	Davies et al. 1983
p754	DXS84	Xp21.1-Xp21.2	PL Pearson, Leiden	Hofker et al. 1985
L1.28	DXS7	Xp11.3	PL Pearson, Leiden	Bakker et al. 1983
p58.1	DXS14	Xp11-cen	L Kunkel, Boston	Kidd et al. 1989
pDP34	DXYS1	Xq13-Xq21.1	DC Page, Cambridge MA	Kidd et al. 1989
A13-R1	DXS87	Xq21-q24	ATCC	Kidd et al. 1989
p43-15	DXS42	Xq24	ATCC	Kidd et al. 1989
pX45d	DXS100	Xq25	BN White, Kingston	Mulligan et al. 1985
pHPT30	HPRT	Xq26	CT Caskey, Houston	Kidd et al. 1989
52A	DXS51	Xq26.2-q26.3	*	Kidd et al. 1989
cX38.1	DXS102	Xq26.2-q27.1	*	Kidd et al. 1989
pVIII	F9	Xq26.3-q27.1	G Brownlee, Oxford	Kidd et al. 1989
cX55.7	DXS105	Xq27.1-q27.2	MH Hofker, Leiden	Hofker et al. 1987
4D-8	DXS98	Xq27.2	RL Nussbaum, Houston	Boggs & Nussbaum, 1984
RN1	DXS369	Xq27.2-q27.3	BA Oostra, Rotterdam	Oostra BA et al. 1990
pc2S15	IDS	Xq28	PJ Wilson, Adelaide	Wilson PJ et al. 1990
U6.2	DXS304	Xq28	N Dahl, Uppsala	Dahl et al. 1989
1A1.1	DXS374	Xq28	KE Davies, Oxford	Patterson et al. 1989
p114.12	F8C	Xq28	*	Kidd et al. 1989
pKSB	G6PD	Xq28	*	Kidd et al. 1989
MN12	DXS33	Xq28	*	Patterson et al. 1987a
St14-1	DXS52	Xq28	JL Mandel, Strasbourg	Oberle et al. 1985

The probes are listed in order down the X chromosome (Keats et al. 1989; Mandel et al. 1989). In chapter 4 data is presented regarding the loci indicated (*). Data regarding these loci were either derived from the literature or provided by a collaborator. (ATCC = American Type Culture Collection).

the size markers (Ogden & Adams 1987). Ethidium bromide is mutagenic and appropriate precautions were taken. The DNA was transferred from the gel to a nylon filter (Gene Screen Plus, Dupont) by Southern blotting (Southern 1975).

DNA probes were radio-labelled by random primer extension (Feinberg & Vogelstein 1983) or nick translocation (Rigby et al. 1977) to incorporate ^{32}P-dCTP (Amersham kits and ^{32}P-dCTP [1mCi/ml]). The VK probes contained repeated DNA sequences. To reduce the non-specific hybridization of these radio-labelled sequences, the VK probes were pre-reassociated with an excess of unlabelled human placental DNA (Sealey et al. 1985) after radio-labelling.

Pre-reassociation of labelled DNA probes. 25 mg of human placental DNA (Sigma) was dissolved in 20 ml of 10 mM TrisHCl/ 0.1 mM EDTA, stood in ice, and sonicated for 30 seconds. The resulting DNA fragments were approximately a kilobase long. The DNA was boiled gently for 15 min. in a microwave oven, cooled, and precipitated with ethanol. The DNA was resuspended in 10 mM TrisHCl/ 0.1 mM EDTA at a concentration of 20 mg/ml.

100 ng of radiolabelled probe was made up to a volume of 25 ul with 10 mM TrisHCL/ 0.1 mM EDTA; 50 ul of the sonicated DNA and 25 ul of 20xSSC (3M NaCl/ 0.45 M Na citrate) were added, mixed, and the solution boiled for 10 min. The solution was chilled on ice for 1 min,

placed in a water bath at 65°C for 10 minutes, and added to warm hybridization solution. The radiolabelled mixture was mixed vigorously, allowed to de-gas, and added to a bag containing the prehybridized nylon filter.

In some instances, unincorporated radiolabelled nucleotides were removed from the DNA probe with GeneClean (Bio 101 Inc; Vogelstein & Gillespie 1979)[2]. All ^{32}-P radiation work was performed according to the appropriate South Australian code of radiation safety.

Nylon filters were wetted with 5xSSC and then prehybridized with 5xSSC/ 50% formamide/ 1% SDS/ 70% dextran/ 0.1 mg/ml sonicated denatured salmon sperm DNA (Sigma) at 42°C for 1 hour (protocol developed by Dr PV Nelson, Adelaide). The radio-labelled probe was heat-denatured (or pre-reassociated) and added directly to the prehybridization solution. The nylon filters were hybridized at 42°C overnight with gentle agitation. They were then washed in 2xSSC/ 0.5% SDS and then in 0.1xSSC/ 0.1% SDS for 30 minutes each at 65°C. The washed filters were exposed to X-Omat film (Kodak) (with an intensifying screen) at -70°C for 1-14 days. The nylon

---

[2] 2-3 volumes of saturated NaI was added to the DNA solution. 5 ul of well-mixed GlassMilk was added, mixed, and allowed to stand for 5 minutes in which time the DNA bound to the glass beads. The beads were washed well with an excess of cold NaCl/ ethanol/ water, and the DNA eluted with 10 mM TrisHCl/ 0.1 mM EDTA at 65°C.

filters were then stripped of the radiolabelled probe by washing in 0.4 M NaOH for 30 min at 42°C, and then in 0.2 M TrisHCl/ 0.1% SDS/ 0.1xSSC for 30 min at 42°C.

SUBCLONING OF VK PROBES

Pre-reassociation of the VK probes was effective in reducing the non-specific hybridization of repeated sequences when probing the somatic cell hybrid lines. However it was less effective when probing lymphocyte DNA, and was not reliable when screening many DNA samples for RFLPs. Therefore, repeat-free fragments of the VK probes which detected loci close to FRAXA were subcloned to allow rapid screening for RFLPs.

Fragments of the VK clones that contained unique DNA sequences had been identified by Dr VJ Hyland (Adelaide, S.A.). Repeat-free fragments of VK21 and VK23 were subcloned into the multiple cloning sites of the plasmids pBR328 (Soberon et al. 1980) or pUC19 (Vieira & Messing 1982). One probe remote from FRAXA, VK17, was subcloned into pUC19 (Suthers et al. 1988c [Appendix D]). VK16 and VK18 were subcloned by J Nancarrow (Adelaide, S.A.). To subclone a specific repeat-free fragment, the intact VK probe was digested with the appropriate combination of enzymes and all the fragments were cloned into a cut de-phosphorylated plasmid. Mini-preparations of DNA were made from 12-24

subclones. The subclone containing the insert of the correct size was identified by digesting the plasmid DNA with the appropriate restriction endonuclease and determining the size of the inserts by gel electrophoresis. All recombinant DNA work was carried out in accordance with the code of practice required by the National Health and Medical Research Council.

Subcloning of DNA fragments in plasmid vectors[3]. *25 ug of the plasmid was digested overnight with a selected restriction endonuclease (EcoRI, HindIII, or SalI), singly or in combination, using the specified buffer. The reaction mixture was then made up to 50 mM TrisHCl/ 0.1 mM EDTA and incubated with 500 units (25 ul) of calf alkaline phosphatase (Boehringer Mannheim) at 37°C for 1 hour. Proteinase K was added to a concentration of 0.1 mg/ml, and the mixture incubated at 37°C for 1 hour. The DNA was extracted with phenol/chloroform, precipitated with ethanol, and resuspended in 25 ul of 10 mM TrisHCl/ 0.1 mM EDTA.*

1 ug each of the prepared vector and the digested DNA fragments were heated separately at 65°C for 10 minutes, and then mixed with ligase buffer (IBI) in a total volume of 200 ul. 1-2 units of T_4 ligase (IBI) were added, and the mixture incubated at 4°C overnight. The

---

3 *This protocol is a hybrid of the protocol given in the IBI product handbook (1986), local knowledge from Dr VJ Hyland (Adelaide, S.A.), and the literature summarized in Greene and Guarente (1987).*

ligation reaction was stopped with EDTA at a concentration of 25 mM.

Strains of E. coli (strain JM83 for pUC vectors; NEK for pBR vectors) were made competent for transformation using the protocol of Hanahan (1983). Cells were grown in liquid media to an optical density (OD_{650}) of 0.2. The cells were chilled on ice for 20 minutes, pelleted, and washed in 0.1 M $MgSO_4$, washed in 0.1 M $CaCl_2$, suspended in 0.1M $CaCl_2$/ 20% glycerol, and stored in aliquots of 200 ul at -70°C.

Prior to transformation, the cells were thawed on ice. The DNA ligation mixture was warmed at 65°C for 10 minutes, snap chilled, and 20 ul was added to the cells. After 30 minutes on ice, the cells were 'heat-shocked' at 42°C for 2 minutes, chilled on ice, and then added to 1 ml of Luria broth. The cells were incubated for 3 hours at 37°C, pelleted, and smeared onto an agar plate containing ampicillin (50 ug/ml). Aliquots of cells were also transformed with the intact plasmid or with the unligated plasmid preparation to act as positive and negative controls respectively for the transformation.

Cell colonies containing pUC vectors with inserts were selected on the basis of ampicillin resistance and blue color change in the presence of X-gal (Jendrisak et al.

1987) added to the agar. Cell colonies containing pBR vectors with inserts were selected on the basis of ampicillin resistance and tetracycline (12.5 ug/ml) sensitivity.

Rapid small scale preparations of plasmid DNA were obtained using the 'Gene Clean' method (Bio 101, Inc.). A bacterial colony was smeared on to a quarter of a 10 cm agar/ampicillin plate, and incubated at 37°C overnight. The cells were scraped off the agar and washed in 10 mM TrisHCl/ 0.1 mM EDTA. The pellet of cells was incubated on ice in a mixture of 8% sucrose/ 5% Triton X-100/ 50 mM TrisHCl/ 5mM EDTA and 0.5 mg of lysozyme. The tube was then placed in boiling water for 2 minutes and spun at 10000G at 4°C for 15 minutes. The supernatant was removed and the DNA extracted using the Gene Clean procedure.

IDENTIFICATION OF RFLPs

There are many restriction endonucleases that could be used when searching for an RFLP at a particular locus. Some endonucleases have recognition sites that are more prone to polymorphic variation than others (Cooper & Schmidtke 1984), and it would be more efficient to initiate a search for RFLPs with these enzymes. Theoretical models for predicting the

relative efficiencies of different enzymes for detecting RFLPs have been suggested (Bishop et al. 1983; Wijsman 1984). The predicted ordering of restriction endonucleases in terms of their relative efficiency is similar to that observed in practice (Devor 1988).

The search for an RFLP was initiated with the 12 enzymes most likely to detect an RFLP i.e. TaqI, MspI, EcoRI, HindIII, PvuII, BglII, PstI, BamHI, Sau3A, SacI, HincII, and RsaI. For each enzyme, DNA samples from three normal women and three fragile X carriers were pooled in one gel track (12 X chromosomes in all), and DNA from a normal male was placed in an adjacent track. Any difference in the pattern of genomic DNA fragments in the pooled versus the hemizygous sample indicated a possible RFLP. As only two lanes were used for each enzyme it was possible to screen for RFLPs with 12 enzymes using just two 12-lane gels. This method has been described (Rousseau et al. 1990). With a total of 13 X chromosomes being surveyed with each enzyme, there was an 85% probability of detecting a two-allele RFLP with a rare allele frequency of 0.15 or greater (Aldridge et al. 1984).

The size of genomic DNA fragments detected by DNA probes was estimated with the program **DNASIZE** (listed in Appendix A). This program estimated the size of an unknown DNA fragment by comparing its position in an agarose gel with the positions of standard DNA size markers.

LINKAGE ANALYSIS - Implementation

The LINKAGE programs (Version 5)(Lathrop et al. 1984, 1985; Lathrop & Lalouel 1988) were generously provided by Dr Jurg Ott of Columbia University. The programs are written in Turbo-Pascal Version 5 (Borland International Inc.) and run on an IBM-AT compatible personal computer. The steps involved in establishing these programs are detailed below.

Computer hardware and Operating System

The computer was a NEC Powermate IV operating under MS-DOS. The computer had 640K of core memory, a 20 Mb hard disc, 80287 numerical co-processor, and 512K of extended memory. The operating system (DOS) limited the amount of memory that could be used by the programs. Insufficient memory occasionally caused a LINKAGE program to crash while running with the error message "Heap/stack overflow" or "Insufficient memory" (or similar). This problem was overcome in 4 ways:

1. Unnecessary programs were removed from memory.
2. The number of pedigrees being studied was reduced to a minimum, or the linkage study was broken into parts (LINKMAP results from separate runs can be combined subsequently; see Chapter 1).
3. All real variables were reduced from 'extended precision' to 'double precision' or to 'single

precision' in the SWITCHES.PAS file; this reduced the memory required to store real numbers but the accuracy of the analysis was reduced.

4. 512K of extended memory was installed in the computer. The LINKAGE programs were then compiled to overlay files as detailed in Ott's notes for LINKAGE Version 5.03. The overlay files were placed on a RAMDISC in the extended memory as detailed in the DOS manual. (It would be possible to install some enhanced memory with an EMS driver. This would provide even more memory for the LINKAGE programs but is reported to be slow.)

Programs

The LINKAGE package consists of 9 programs for analyzing general pedigrees of arbitrary structure. There are also modified fast versions of these programs for analyzing the CEPH pedigrees.

The programs for use with general pedigrees are:

1. PREPLINK. This was used to create a datafile which contained information about the loci being studied.

2. PEDPOINT. This modified the list of pedigree data and genotypes in the infile and wrote a new pedfile. The program MAKEPED was supplied and was

meant to do the same task, but it was not reliable.

3. UNKNOWN took the datafile and pedfile, checked for data inconsistencies, and inferred unknown genotypes.
4. LODSCORE took the output of UNKNOWN and calculated two-point lodscores between pairs of loci. LODSCORE was used to check the recombination fractions between all pairs of loci before moving on to more complex analyses.
5. LINKMAP took the output from UNKNOWN and estimated the most likely position of an unknown locus on a known genetic map. This was the most appropriate program to use for localizing a disease gene on the human gene map.
6. ILINK took the output from UNKNOWN and calculated the likelihood of a given locus order. It was used to determine the probable order and recombination fractions for a number of unknown loci.
7. MLINK was used to calculate the lod score for a given recombination fraction between two loci (e.g. for preparing a LOD score table) and for risk analysis.

8. LINKAGE CONTROL PROGRAM (LCP) provided an easy environment for using programs 3-7.

9. LINKAGE REPORT PROGRAM (LRP) took the rather obscure output from the LINKAGE programs and summarized it in more comprehensible form.

When analyzing the CEPH pedigrees, the CEPH data base programs effectively did the work of PREPLINK and PEDPOINT. Unfortunately, the datafile prepared by the CEPH data base programs was not correct: it correctly specified that the mutation rate in males was zero[4], but did not specify the female mutation rate. This value (0.0) was added manually to line 2 of the data file. The programs CFACTOR and CINFER replaced the program UNKNOWN for CEPH analyses. The programs CLODSCORE, CMAP, and CILINK replaced LODSCORE, LINKMAP, and ILINK respectively for CEPH analyses. The Linkage Control Program and Linkage Report Program controlled all the CEPH analysis programs.

Compilation

The LINKAGE programs were supplied in both compiled and non-compiled versions. In view of the computer memory constraints and the size of the analyses done during this project, it was necessary to compile the programs specifically for each analysis. The compiler used was Turbo-Pascal Version 5 (Borland International). This compiler was presented in two formats. One format provided a 'user-friendly' screen of

---

4 The CEPH analysis programs are designed to analyze the linkage relationships of co-dominant loci with no mutation in either sex.

menus and advice. The second format (TPC: a command-line compiler) was more cryptic but used less memory. It was possible to compile larger programs with the command-line compiler than with the 'user-friendly' compiler. All the LINKAGE programs had to be compiled with the command-line compiler .

In compiling the LINKAGE programs for general pedigrees, two files needed to be modified, SWITCHES.PAS and LINKAGEC.PAS. The various options available in these two files are detailed in Ott's notes to Version 5.03.

SWITCHES.PAS was used in compiling the 3 main LINKAGE programs (5-7 above). For SWITCHES.PAS the options were set as follows:

```
{$DEFINE linkmap}   {Program name;  mlink or linkmap or ilink}
{$DEFINE double}    {Precision; single, double, or extended}
{$O-}               {Overlays; + or -}
{$N+}               {Use numeric coprocessor; + or -}
{$E-}               {Emulate coprocessor; + or -}
```

In Turbo Pascal, brackets { } are used to indicate either a comment or a critical command for the compiler. { DEFINE .....} is a comment for the programmer's benefit; {$DEFINE .....} is a compiler directive. If the '$' was in the wrong place the programs wouldn't compile correctly. The program to

be compiled had to be identified in the first line. If the program was to be compiled to overlay files (i.e. {$O+}) the overlay path in the corresponding program (e.g. LINKMAP.PAS) had to be altered to read:

```
ovrinit('E:\LINKMAP.OVR');
```

where E was the disc name for the RAMDISC.

LINKAGEC.PAS was used to compile UNKNOWN in addition to the three main programs. It listed the constants that would be used by the programs. Ideally the constants should be set as low as possible. Many of these constants are used to determine the size of arrays used by the programs. If a constant was set too low, the LINKAGE programs crashed with the error message "Array subscript out of range" (or similar). Sometimes it was possible to use the Turbo Pascal compiler to determine which array constant was too small. However it was difficult to track errors due to low values of MAXNEED or MAXCENSOR. If the constants were set too high there was insufficient memory to compile the program. (Turbo Pascal limits the amount of memory available to all arrays in a program; this is a Turbo Pascal limitation, not a DOS problem, and cannot be resolved by adding more memory). In practice, if using LINKMAP the analysis was limited to about 6 two-allele loci in approximately 10 families.

The following values were initially used for the constants:

```
maxneed=2000;        {MAXIMUM NUMBER OF RECOMBINATION
```

```
maxcensor=500;      {MAXIMUM FOR CENSORING ARRAY}
maxloop=1;          {MAXIMUM NUMBER OF LOOPS; must be >0}
scale=1.0;          {SCALE FACTOR}
scalemult=2.0;      {SCALE WEIGHT FOR EACH LOCUS INCLUDED}
fitmodel=FALSE;     {TRUE IF ESTIMATING PARAMETERS OTHER THAN

dostream=TRUE;      {STREAM FILE OUTPUT; essential for LCP &
LRP} byfamily=TRUE ;    {GIVE LIKELIHOODS BY FAMILY}
feedback=TRUE;      {DESCRIPTION OF PATH THROUGH FAMILIES;
                    indicated where the program was in the
                    analysis, and showed if the program was
                    stuck at some error. However, feedback
                    slowed the analysis }
findlost=FALSE;
```

When compiling the CEPH analysis programs the files SWTHG.pas and THGC.pas replaced SWITCHES and LINKAGEC respectively. The SWTHG file contained a special compiler directive that indicated whether the analysis would deal with X-linked or autosomal data. The CEPH programs were much more efficient in the use of memory, and could handle many more alleles and loci than the general programs.

Coding data

Ott has outlined the coding conventions in his Version 5.03 documentation and in the more extensive documentation that accompanied earlier versions of LINKAGE.

There are a number of ways of defining loci for LINKAGE analysis. A co-dominant RFLP gives information about a person's genotype at a locus. There are two ways of defining genotypes for LINKAGE analysis: either as a series of 1's and 0's (binary factor system) or as allele numbers. Either system may be used but it was essential that the same system be used for the RFLP in the infile and the datafile. The binary factor system was used for coding RFLPs in both the CEPH and fragile X pedigrees.

There are four ways of defining a disease locus. If a genotype can be confidently inferred from the phenotype (i.e. no mutation, no homozygous affected people, and penetrance of 1.0) then the disease locus could be coded as a genotype using binary factors or allele numbers. However, in the fragile X syndrome it is usually not possible to be certain about a person's genotype if they are phenotypically normal. It is possible to code a disease locus as a phenotype and let the LINKAGE programs infer the probable genotype/s and incorporate that in the analysis. Phenotypes can be coded as affection status or as a quantitative trait. Coding the fragile X as affection status was the appropriate method as it was possible to make allowance for the penetrance of the condition. It is also possible to specify different penetrance values for different groups of people (described as different liability classes). This is important in risk

analyses in the fragile X syndrome (see below) but was not utilized in mapping studies of FRAXA.

Using LINKMAP and CMAP

This was the most appropriate program to use when mapping a new locus on a known genetic map. The background map had to be reliable and to have been generated using different data[5]. Published maps and the output from the Linkage Report Program give the distance between loci in cM (centimorgans). However the linkage programs do all the calculations with recombination fractions, and the Linkage Control Program requests the recombination fractions between loci. As mentioned in Chapter 1, there are a number of mapping functions that describe the relationship between genetic distance and recombination fractions, and they are not interchangeable. The various functions make different assumptions about interference and inconsistent use would give wrong results (Keats et al. 1989). LINKMAP made no allowance for interference, and Haldane's mapping function (which also makes no allowance for interference) was the most appropriate function to use. The program **MORGAN** was written

---

5 The precision of a localization determined by LINKMAP is to a large extent governed by the accuracy of the background genetic map (Suthers & Wilson 1990). If there is little independent evidence to support the background map, the localization determined by LINKMAP may be far less precise than it appears. If there is not a reliable background genetic map available, it would be more appropriate to use the program ILINK which makes no assumptions about the positions of nearby loci.

for converting recombination fractions to genetic distances using the mapping functions of Haldane and Kosambi (Appendix A).

The output from LINKMAP consists of a series of paired data-points (genetic location and location score) that must be plotted for interpretation. A simple plotting program, **PLOT** (Appendix A), was written to plot these points on the computer screen, thus speeding evaluation of the results.

Using ILINK and CILINK

ILINK and CILINK calculated the likelihood of a given order of loci and the best estimates of the recombination fractions. The likelihood was expressed either as a transformed likelihood (-2ln[likelihood]) or as Ott's Generalized LOD Score. The difference in the likelihood of two genetic scenarios (e.g. different orders of loci) was 10 raised to the power of the difference in the Generalized LOD Score values. It was usually not possible to give a p-value for the difference in likelihoods. However, if the two scenarios being compared had

the same order of loci and one scenario could be 'nested'[6] in the other, then the difference in the two transformed likelihood values [-2ln(likelihood)] could be treated as a X^2 value. The degrees of freedom is the difference in the number of parameters allowed to vary in the two models (Mendenhall et al. 1986, p. 431). A fixed model (e.g. all recombination fractions specified) has no varying parameters. An example of this is given in Chapter 7.

The Linkage Report Program interpreted the output of ILINK or CILINK, indicated if the analysis was valid, and calculated the relative likelihoods of different locus orders. If the analysis was not valid, Linkage Report Program did not indicate why, nor how the problem might be rectified. The file FINAL.out listed the 'exit condition' for each run of the program:

> *The cryptic message 'Specified tolerance on normalized gradient met' meant that the iterative process at the heart of the program was completed successfully.*

---

6 *That is, in comparing two models the models must have the same order of loci and one model must contain all the variable and fixed parameters (i.e. recombination fractions, male:female recombination ratio etc.) of the other. For example, if a recombination fraction is estimated in one model and fixed in the other, the fixed parameter is considered as 'nested' within the estimated parameter as the estimated value could equal the fixed value. Hence, the difference in likelihoods of the two models may be given a p-value. However, if the recombination fraction is fixed at different values in both models, the two parameters are not nested, and the two models can only be compared in terms of relative likelihoods.*

The message 'Maximum possible accuracy reached' indicated that the precision of the real numbers was not sufficient to allow a conclusion to be drawn. If the programs were re-compiled with increased precision of real numbers (as detailed above), a repeat analysis was often successful.

Before running ILINK or CILINK, it was necessary to specify initial values of the recombination fractions. If the initial values were too removed from the true values, an accumulation of rounding error prevented accurate estimation of the recombination fractions. The program would halt at that point, and list the current estimates of the recombination fractions and the message 'Accumulation of rounding error prevents further progress'. If the program was rerun using those estimates of the recombination fractions as initial values, the analysis could be completed successfully.

ILINK and CILINK did not support the inclusion of interference in the analysis if there were more than three loci. Every multipoint analysis performed during this project involved data at four or more loci, and interference was not considered.

LINKAGE ANALYSIS - Normal (CEPH) Pedigrees

In addition to providing DNA samples from 40 three-generation pedigrees, CEPH provided a database of genotypes at over 800 loci and associated software (Dausset et al. 1990). The computer programs provided by CEPH were written in the language 'C', and did not require modification.

In the course of this project, the families were genotyped at three additional polymorphic loci, DXS296 (VK21), DXS297 (VK23), and IDS (pc2S15). The genotyping of the CEPH pedigrees was performed by HM Kozman and J McCure under the candidate's direction. As the database did not contain a complete listing of genotypes at other loci near FRAXA, further CEPH genotypes at five loci near FRAXA were obtained from Dr I Oberle (Strasbourg).

Two programs were written to minimize data entry errors and to aid analysis of this data (Appendix A). The program **TEMPLATE** provided standard listings of pedigree numbers, individual identification numbers, and DNA sample numbers for use when making Southern blots of CEPH pedigree DNA. This order was stored in the computer so that new genotypes could be entered into the database in the same order in which they were read from the autoradiograph.

Some of the polymorphic loci in the CEPH database could be defined by a number of RFLPs i.e. one DNA probe would detect different RFLPs in genomic DNA samples digested with different restriction endonucleases. In genetic terms, these RFLPs are at the same location, and linkage analysis of these loci would be simplified if all the RFLP data at a locus could be condensed or merged to a single pair of alleles for each individual. CEPH provided a program HAPLO that was designed to perform this condensing operation. However, this program was found to be unreliable. Five of the nine polymorphic loci mapped in the CEPH pedigrees (Chapter 5) had more than one RFLP. The RFLP data at these loci was condensed by hand prior to the analysis.

The CEPH software also included a program for drawing simple three-generation pedigrees with specified genotypes listed under each individual. The phase of the alleles was not indicated, and it was difficult to check the data and identify recombination events by hand. The program **XPHASE** was written to perform two functions. It checked the data to ensure that there were no inconsistent genotypes. The program then examined the paternal and maternal grandfather's genotypes and inferred the phase of X-linked loci throughout the pedigree. These phase-known data could then be printed using the CEPH pedigree-plotting program, and recombination events could be readily identified (Appendix B).

LINKAGE ANALYSIS - Fragile X pedigrees

The 35 fragile X pedigrees had been genotyped by Dr J Mulley at the polymorphic loci F9, DXS105, DXS98, and DXS52. These pedigrees were genotyped at three additional polymorphic loci, DXS296 (VK21), DXS297 (VK23), and IDS (pc2S15), by the candidate. As affected individuals with no affected relatives could represent new mutations (Sherman et al. 1988), pedigrees with a single affected individual were excluded from the analysis.

To minimize errors during data entry and manipulation, all genotypes were entered into a computerized database. Fragile X pedigrees were plotted with genotypes listed below each individual by combining this database with a general purpose pedigree plotting program, TEXTPED (provided by Dr M Badzioch, Houston)(Appendix C).

The genotypes at DXS296 and nearby loci in a further 77 fragile X pedigrees were obtained on a collaborative basis from other centers. The collaboration was initiated and arranged by the candidate. The basis on which the pedigrees were selected is presented in Chapter 6. The collaborators and the number of pedigrees provided by each group are listed in Appendix C.

In setting parameters such as FRAXA allele frequency and mutation rate, the data of Sherman et al. (1985, 1988) were chosen. The allele frequency was 0.0006, and the mutation rate was 0.00024 in males and 0.00048 in females. With regard to penetrance figures, two strategies have been used. The first strategy was to use a single penetrance figure for each sex which ignores maternal intellectual status. As detailed in Chapter 1, the penetrance figures for mental retardation are 0.79 in males and 0.37 in females in this situation (Sherman et al. 1985). However carriers may express the fragile X without being mentally retarded. More carriers would be identified if the penetrance figures were based on the presence of mental retardation or fragile X expression. If the maternal intellectual status is ignored, these penetrance values are 0.79 in males and 0.56 in females (Mulley & Sutherland 1987).

The second approach has been to specify different penetrance figures according to the mother's intellectual status (Mulley & Sutherland 1987). The critical difference in these two approaches is that the second approach specifies a penetrance of 1.0 for the sons of mentally retarded females. In linkage studies involving many fragile X pedigrees from various sources there is a danger of incorrectly specifying a carrier as retarded. If this happens, the penetrance among her sons will be incorrectly high and the recombination fraction will be over-estimated (Clerget-Darpoux et al. 1986). For this

reason the intellectual status of the mother was ignored when specifying the penetrance parameters. If pedigree data indicated that an apparently normal individual of either sex was an obligate carrier, that individual was coded as affected for the linkage analysis.

It was not possible to analyze all the fragile X pedigrees simultaneously with the program LINKMAP. As mentioned above, LINKMAP was effectively limited to analyzing 6 loci (i.e. FRAXA and 5 RFLPs) in 5-10 pedigrees. To circumvent this limitation, each family was analyzed separately with LINKMAP using the same background genetic map. (The development of the genetic map is presented in Chapter 5). The most likely position of FRAXA on the genetic map was noted for each family. The individual sets of multipoint LOD scores were summed at the corresponding points on the genetic map to provide an overall set of multipoint LOD scores for the location of FRAXA. This manipulation of data was managed by a series of small programs (written by the candidate) which extracted, tabulated, and summed the relevant figures.

The advantage of this approach was that the contribution of each family to the location of FRAXA could be examined, and atypical results be checked. It was also possible to extend the analysis to six RFLPs. Two polymorphic loci, DXS369 and DXS297, were located at the same position on the genetic map (Chapter 5). None of the fragile families used in this

multipoint linkage study of FRAXA location had been genotyped for RFLPs at both of these loci. In those families informative for the polymorphism at one locus, the other locus was excluded from the analysis. No information was lost by this manoeuvre. On the contrary, the inclusion of another locus in the analysis increased the information available for the multipoint analysis (Lathrop et al. 1985). There were 101 families that were informative at one or more of these six polymorphic loci, and were included in the analysis.

HOMOGENEITY TESTS

Homogeneity of two-point recombination fractions in both the CEPH and fragile X pedigrees was assessed using the HOMOG package of programs (Ott 1985). One of the programs in this package, MTEST, is an implementation of Morton's homogeneity test (Chapter 1)(Morton 1956). The rationale of Morton's test is as follows.

> *A linkage study is usually performed under the assumption of linkage homogeneity, i.e. the recombination fraction in all the families being studied is the same. The likelihood (expressed as log_{10}) of there being one recombination fraction in all the families is simply the peak LOD score in the pooled data. However, in any linkage study the peak LOD score will occur at a different recombination fraction in each*

family studied. This reflects the different pedigree structures, numbers of informative individuals in each family, and the stochastic nature of recombination. The likelihood of each family having its own unique recombination fraction is equal to the sum of the peak LOD scores in each family. If there is linkage homogeneity in the sample of families being studied, the two likelihoods will not be significantly different. If the difference in the two likelihoods is transformed to be -2ln(likelihood difference), the transformed value may be treated as a X^2 value and its significance determined. The number of degrees of freedom is one less than the number of families.

Morton's test is very conservative because small families add little to the total likelihood values but do increase the degrees of freedom. In practice, it may be difficult to assign reliable significance values to the difference in the likelihoods (Ott 1985, p.113; Risch 1988). The test is more valuable when comparing groups of families that can be divided prior to the linkage analysis (Chapter 1). This constitutes the 'pre-divided sample' test of homogeneity. The two likelihoods are calculated between the groups rather than between the families, and the degrees of freedom of the transformed likelihood difference is one less than the number of groups.

MTEST was used to determine whether there was a significant difference in the recombination fractions between two loci in the CEPH pedigrees versus the fragile X pedigrees (Chapter 7).

When testing for heterogeneity of two-point recombination fractions in the fragile X pedigrees, implementations of Ott's A-test (Ott 1985) were used. The choice of program was determined by the hypothesis being tested. Ott has distributed a variety of HOMOG programs, each of which evaluates a different alternative hypothesis e.g. is the recombination fraction in males different to that in females; do some families demonstrate linkage and others not; can the families be divided into four groups on the basis of the recombination fractions. To avoid multiple comparisons, the choice of alternative hypothesis must be made prior to the analysis. As Brown et al. (1987) had suggested that fragile X families could be divided into two groups on the basis of the F9:FRAXA recombination fraction, the program HOMOG2 was used. The HOMOG2 program uses maximum likelihood estimation to evaluate three hypotheses. The null hypothesis (H_0) is that the two loci being examined are unlinked in all the families (θ=0.50). The next hypothesis (H_1) is that the two loci are linked at the same recombination fraction in all the families (θ <0.50). The final hypothesis (H_2) is that the families

comprise two groups, with the loci linked at different recombination fractions ($\theta_1 < \theta_2 < 0.50$).

The transformed values of the three likelihoods (-2ln(likelihood)) may be compared using the X^2 distribution with one (H_1 vs H_0) or two (H_2 vs. H_1) degrees of freedom. Under these conditions, the tests are generally conservative.

The programs HOMOG2 and HOMOG3 were used to perform multipoint tests of linkage homogeneity. In the multipoint linkage analysis of fragile X families (Chapter 6), the likelihood (expressed as a multipoint LOD score) of FRAXA being located at specific points along a predefined genetic map had been calculated for each family using the program LINKMAP. The background genetic map used for this analysis had been determined independently in the CEPH families (Chapter 5), and consisted of the loci DXS98, DXS369, VK23B, VK21, IDS, and DXS304. For homogeneity testing, multipoint LOD scores were calculated at 10 locations in the interval VK23B:VK21 and at 5 locations in each of the other intervals.

In analyzing these data with a multipoint linkage homogeneity test, the null hypothesis (H_0) was that FRAXA was not located on the genetic map; H_1 was that there was a single location for FRAXA on the map; H_2 was that the families could be divided into two or three groups (depending on the program

used) with different FRAXA locations. When performing a multipoint test of linkage homogeneity, the differences in the likelihoods do not have a defined distribution (Clayton et al. 1988; J. Ott, personal communication), and significance values cannot be assigned to differences in the likelihoods of the three hypotheses.

ESTIMATION OF CONFIDENCE INTERVALS IN LINKMAP ANALYSES

The program LINKMAP usually generates a complex likelihood function for gene location, and this function may have two or more maxima. As detailed in Chapter 1, there is no standard statistical technique for determining a confidence interval for gene location in such a situation (Lathrop et al.1984). The "one-LOD-unit-down" method is often applied in multipoint linkage analysis but the significance level of such a confidence interval is unclear (Keats et al. 1989).

Computer-intensive resampling (or 'bootstrap') techniques are becoming widely used in situations where complex problems elude formal analytical solution (Efron & Tibshirani 1986). Wilson and La Scala (1989) have used the resampling methodology for determining confidence intervals for the recombination fractions between a disease locus and marker loci, and for determining the evidence for locus order for data from a set of nuclear families.

The application of the resampling method for determining an approximate confidence interval for disease gene location on a known genetic map is described below. This method was developed in association with Dr SR Wilson (A.N.U., Canberra), and has been published (Suthers & Wilson 1990 [Appendix D]).

Let D represent the unknown true position of the disease locus (with respect to an *a priori* determined origin). One type of resampling procedure for constructing a confidence region for D is as follows. From the observed data we have an estimate of D, $\hat{d}^*$. The essence of the resampling approach is to take a random resample from our original data and to repeat the estimation procedure with this resample to obtain $\hat{d}_1^*$. This resampling procedure is repeated B times to give $\hat{d}_1^*, \hat{d}_2^*, \ldots, \hat{d}_B^*$. The simplest method of determining a confidence interval is the 'percentile method' (Efron & Tibshirani 1986;Hinkley 1988; DiCiccio & Romano 1988). The percentile method interval is just the interval between the corresponding percentiles of the resampled distribution of $\hat{d}^*$. If B = 49, the approximate 98% confidence interval for gene location is the range of the values. If B = 1000, the 95% confidence interval is the range of 95% of the locations centered at the median.

There are a number of ways in which the resampling can be performed (Wilson & La Scala 1989). The resampling method was

applied in two situations which each required different resampling strategies. The mental retardation gene MRX1 was mapped to Xp11 (Chapter 3). Here the multipoint linkage data was derived from a single pedigree, and novel resampling strategies had to be employed. In Chapter 6, a collaborative multipoint linkage study of FRAXA is described. This study included data from 101 families, and a simpler resampling strategy was used.

Resampling Strategy in a single pedigree

The major difficulty in applying this method to the multipoint linkage analysis of a single pedigree was to choose an appropriate random resampling protocol. Although each meiosis in the pedigree is an independent event, the information that can be obtained from each meiosis is dependent another complex factors (such as pedigree structure) that enable, say, phase to be inferred. An essential point to keep in mind is that the resampling simulation should, implicitly or explicitly, simulate each component of variability. Two different resampling protocols were used to determine whether the conclusions concerning the disease gene location were robust to the exact form of resampling chosen.

The first protocol (Protocol I) was based on a simulation method described by Lathrop et al. (1987). For this protocol the pedigree structure and genotypes for generations I and II

in each resample were the same as in the original data. If an individual in generation II had children then the original sibship in generation III was randomly resampled to create a sibship of the same size for generation III of the resample. Similarly resampled sibships were added to generation IV of the resample. In this way each resample had the same pedigree structure as the original data set, and the genotypes within each sibship in generations III and IV were randomly selected.

The second resampling protocol (Protocol II) was prompted by the Elston-Stewart algorithm (Elston & Stewart 1971) and exploited the conditional structure of the likelihood formulation. Each individual in the pedigree was regarded as being a branch end and these branch ends were taken to be independent for the resampling. If an individual was chosen in the resample his parents, grandparents etc. up the tree were included. If two or more individuals had a common ancestor then the branches were merged. The resampling stopped once the number of people in the resample (including parents, grandparents etc.) totalled 31 (the pedigree size). Using this protocol the pedigree size was fixed but the pedigree structure varied with each resampling.

With both protocols females of unknown carrier status were included and the carrier status of a woman was determined by whether or not she had any affected sons or grandsons in that

resample. Therefore each resample varied in the number of affected males, normal males, obligate carrier females and females of unknown carrier status.

The number of resamples taken, B, depends on the form of $D-\hat{d}^*$, and will often be at least 100 (Hinkley 1988). For the study described in Chapter 3, 19 resamples were taken under each protocol and then a further 30 resamples were taken. The qualitative conclusions were not changed by the increase in the number of resamples. In view of the considerable computing involved there seemed to be no advantage to further increasing the number of resamples. The resampling of this pedigree was performed by Dr Wilson.

The extent of the confidence interval determined with this approach is critically dependent on the background genetic map used for the LINKMAP analyses. Although the background genetic map is regarded as fixed, in fact there is usually a degree of uncertainty about the precise location of the various points along the map. The importance of having an accurate background genetic map for estimating a confidence interval for gene location is discussed further by Suthers and Wilson (1988; [Appendix D]).

Resampling Strategy with data from many pedigrees

The rationale in resampling with data from many pedigrees is to randomly resample the pedigrees rather than the individual

meioses. In the study reported in Chapter 6, 101 fragile X families were included in the LINKMAP analysis. An approximate 95% confidence interval for FRAXA location was determined as follows.

Multipoint LOD scores were calculated at specific points along the genetic map for each of the 101 pedigrees in the analysis. These sets of LOD scores were randomly resampled (with replacement) 101 times. The resampled LOD scores corresponding to each point on the genetic map were summed. The peak multipoint LOD score indicated the most likely location of FRAXA in that resample. As this strategy did not involve repeated linkage analyses it was very rapid, and this resampling process was repeated 1000 times. The range of 95% of the resampled FRAXA locations (centered at the median location) indicated the approximate 95% confidence interval for FRAXA location.

The program **BOOTMAP** was written by the candidate to perform the repeated resampling and summation of the multipoint LOD scores. This program is listed in Appendix A along with an example of the program output.

Chapter 3

THE LOCUS FOR A NON-SPECIFIC FORM OF X-LINKED MENTAL RETARDATION (MRX1) IS LOCATED AT Xp11.

SUMMARY

Linkage analysis of the locus causing a non-specific form of X-linked mental retardation (MRX1) was performed in a large pedigree. The affected males had moderate mental retardation; in all other clinical respects and cytogenetically they were normal. Linkage analysis was performed with a number of markers on the X chromosome. No recombinants were observed between MRX1 and the locus DXS14 (p58.1) located at Xp11-cen; the peak LOD score was 2.90. Close linkage between MRX1 and loci which flank the fragile X locus, FRAXA, was excluded.

The combination of multipoint linkage analysis and a novel resampling strategy indicated that the approximate 98% confidence interval for MRX1 location was 0 to 9 centiMorgan distal to DXS14. Multipoint risk analyses based on this confidence interval for MRX1 location had a dramatic impact on the genetic advice available to consultands in this family.

Thus a locus responsible for non-specific X-linked mental retardation has been localized to the proximal portion of Xp. This location is genetically distant from FRAXA and indicates that MRX1 is not allelic to FRAXA.

INTRODUCTION

X-linked mental retardation is common, occurring in approximately one in 600 male births (Sutherland & Hecht 1985, p.111). Most affected men do not have any recognised disorder of metabolism, chromosome abnormality, or syndromal anomalies. It has been indirectly estimated that there are 7 to 19 X-linked loci causing non-specific X-linked mental retardation (Morton et al. 1977; Herbst & Miller 1980). The fragile X syndrome accounts for at least one quarter of these cases. The fragile X locus (FRAXA) is located at Xq27.3. The location of other loci causing non-specific X-linked mental retardation is not known. The high mutation rate in the fragile X syndrome and the variable cytogenetic expression of the fragile X raise the possibility that loci causing non-specific X-linked mental retardation without the fragile X could be allelic to FRAXA.

The position of a locus causing a non-specific, cytogenetically normal form of X-linked mental retardation in a large pedigree was determined by linkage analysis with polymorphic loci on the X chromosome. This mental retardation locus has been given the symbol MRX1 (Mandel et al. 1989). The 98% confidence interval for gene location was estimated by using a resampling strategy. On the basis of this localization, genetic risk estimates could be calculated for

women in the family.

METHODS

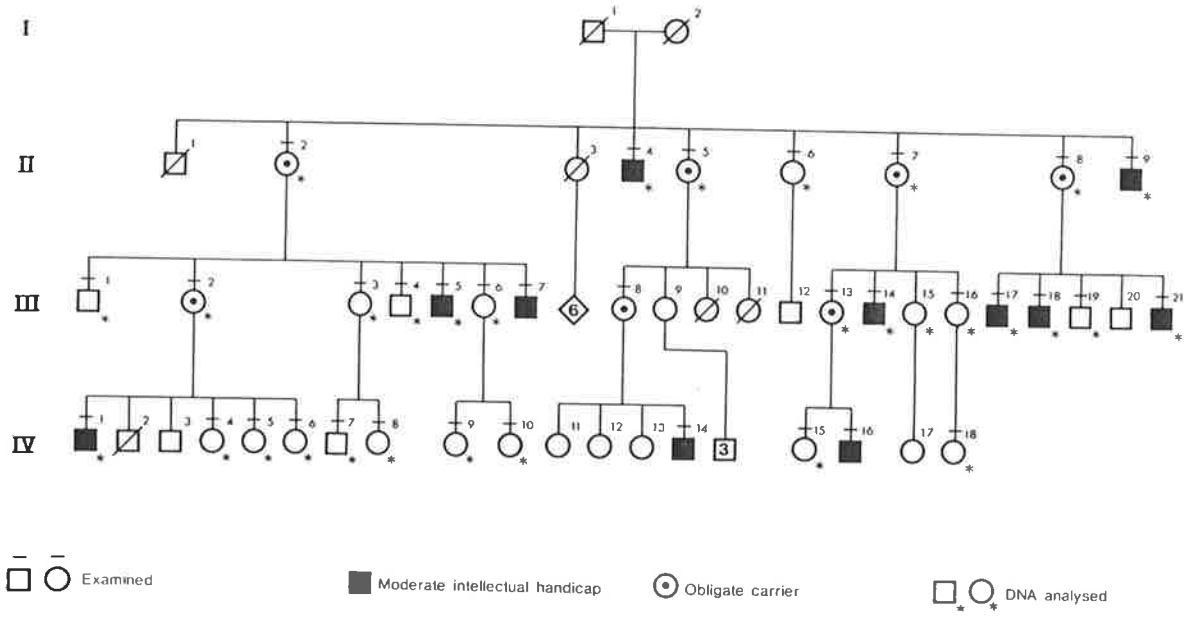
The family was identified and examined by Dr G Turner (Randwick, N.S.W.), and has been described (Turner et al. 1971, family 4). The pedigree is shown in Fig. 3-1, and includes children born since that report. Males with non-specific X-linked mental retardation occurred in each of the three surviving generations. The affected men all had moderate mental retardation. They did not have minor anomalies, were of normal height, and had a normal head circumference. They did not have abnormal neurological signs, seizures requiring medication, or malformations. The karyotypes of the men II-4 and III-5 were normal with no expression of the fragile X in low-folate media.

The obligate carrier women were not overtly intellectually handicapped. Formal psychometric testing was not performed. They had no minor anomalies.

DNA was analyzed from 31 members of the pedigree, including eight affected men (Fig. 3-1). It was not possible to obtain blood for DNA analysis from a further three affected boys (III-7, IV-14 and IV-16). Blood was collected from four normal men and six obligate carrier women. Blood was also collected from the 13 women of unknown carrier status.

Figure 3-1.

Pedigree of the family studied.



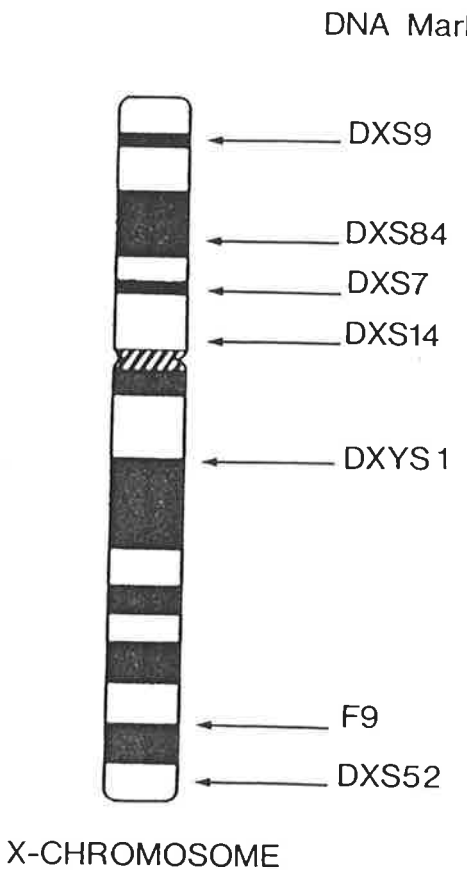
The positions of the loci used in this linkage study are summarized in Figure 3-2. Details of the polymorphisms at these loci have been compiled by Kidd et al. (1989) and are presented in Table 2-C. DNA methods and the linkage analysis programs used are described in Chapter 2. The penetrance of the mutant MRX1 allele was taken to be 1.0, the mutation rate being zero, and the MRX1 allele frequency being 0.0001.

For multipoint linkage analysis the program LINKMAP was used. The MRX1 locus was localized in relation to the three loci, DXS7 (probe L1.28), DXS14 (p58.1), and DXYS1 (pDP34). The order of these loci and the recombination fractions between them were derived from the summary map of the X chromosome presented at the Tenth International Workshop on Human Gene Mapping using Rao's mapping function (Table 24 in Keats et al. 1989). The recombination fractions were 0.16 (DXS7-DXS14) and 0.14 (DXS14-DXYS1). LINKMAP limited the calculation of location scores to discrete points along the genetic map defined by these three loci. (The location score is twice the natural logarithm of the odds for that location of MRX1 on the genetic map versus no linkage.) The points at which location scores were initially calculated are indicated in Figure 3-3.

The approximate 98% confidence interval for MRX1 location was estimated using the resampling method for a single pedigree (Chapter 2). Selected resamples were re-analyzed in greater

Figure 3-2.

The approximate positions of polymorphic loci used in this study are indicated on the X chromosome.



detail using a finer grid of location score calculations around points of interest.

Two-point and multipoint risk estimates for females of unknown carrier status were calculated using the program MLINK (Version 4.6) (Lathrop et al. 1985).

RESULTS

Linkage analysis

The results of two-point linkage analysis are summarized in Table 3-A. No recombinants were observed between MRX1 and the locus DXS14 which is located at Xp11-centromere. Definite recombination was observed between MRX1 and the adjacent loci DXS7 and DXYS1. MRX1 was not linked to two loci which flank FRAXA, F9 and DXS52. The individual genotypes at DXS14 and the flanking loci DXS7 and DXYS1 are listed in Table 3-B.

On the basis of two-point linkage analysis, the regional localization of MRX1 was Xp11.3-Xq21.1. The approximate 90% confidence interval for the recombination fraction between MRX1 and DXS14 was 0-0.20 (Conneally et al. 1985).

The result of multipoint linkage analysis of MRX1 in relation to DXS7, DXS14, and DXYS1 is shown in Fig. 3-3. The peak location score was 18.25 at -5 centiMorgan (cM) relative to DXS14. The discontinuities of the likelihood function at DXS7 and DXYS1 indicated that MRX1 was not located at those loci.

Estimation of a confidence interval for MRX1 location

Applying the "one-LOD-unit-down" method (Conneally et al. 1985; Keats et al. 1989) the confidence interval for MRX1 location was -17 to +5cM from DXS14, i.e. the confidence

Table 3-A.

LOD scores from two-point linkage analysis of MRX1 and polymorphic loci.

	Recombination fractions (θ)								
MRX1 vs.	0.001	0.01	0.05	0.10	0.20	0.30	0.40	LOD_{max}	$\hat{\theta}$
DXS9[a]		-2.99	-1.03	-0.30	0.23	0.35	0.25	0.35	0.30
DXS84[a]		-1.66	0.18	0.76	0.99	0.79	0.41	0.99	0.19
DXS7	-1.24	0.70	1.82	2.06	1.89	1.38	0.68	2.07	0.12
DXS14	2.89	2.85	2.67	2.43	1.93	1.37	0.74	2.90	0.00
DXYS1	-11.27	-6.30	-2.93	-1.61	-0.48	-0.03	0.10	0.10	0.40
F9[a]		-8.50	-4.59	-2.94	-1.45	-0.69	-0.25	0.00	0.50
DXS52[a]		-9.53	-6.20	-4.58	-2.42	-1.20	-0.46	0.00	0.50

Note a Two-point linkage analysis at these loci did not include the normal males III-12 (no DNA collected) and IV-7.

$\hat{\theta}$ recombination fraction at LOD_{max}.

Table 3-B.

Genotypes at DXS7 (L1.28), DXS14 (p58.1), and DXYS1 (pDP34) for 31 members of the pedigree analyzed.

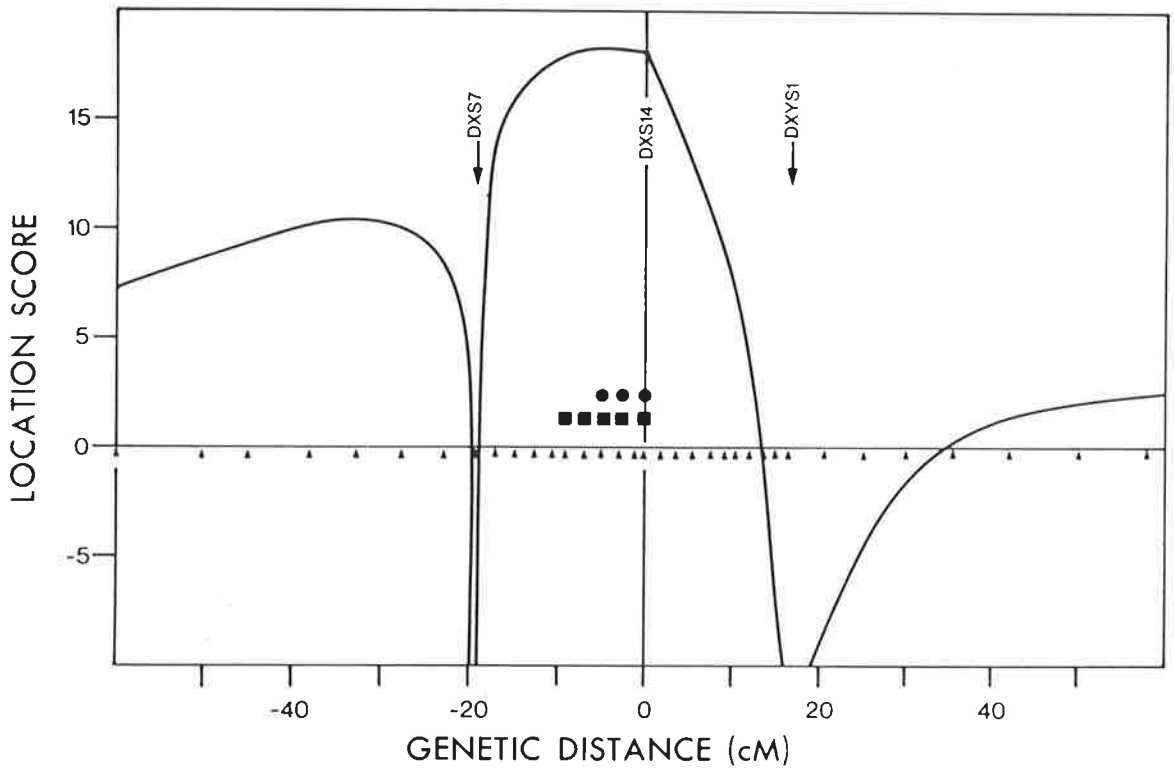
Individual	Status[a]	DXS7	DXS14	DXYS1
II-2	C	A1/A2	A1/A2	A1/A2
II-4	MR	A1	A1	A2
II-5	C	A1/A2	A1/A2	A1/A2
II-6	U	A2	A1/A2	A2
II-7	C	A1/A2	A1/A2	A1/A2
II-8	C	A1/A2	A1/A2	A2/A2
II-9	MR	A1	A1	A2
III-1	N	A2	-	A2
III-2	C	A1/A1	A1/A2	A2/A2
III-3	U	A1/A2	A2	A2
III-4	N	A1	A2	A2
III-5	MR	A1	A1	A1
III-6	U	A1/A2	A2	A2
III-13	C	A1/A2	A1/A1	A2/A2
III-14	MR	A1	A1	A1
III-15	U	A2	A1/A2	A2
III-16	U	A2	A1/A2	A1/A2
III-17	MR	A1	A1	A2
III-18	MR	A1	A1	A2
III-19	N	A2	-	A2
III-21	MR	A2	A1	A2
IV-1	MR	A1	A1	A2
IV-4	U	A1	A1/A2	A2
IV-5	U	A1	A1/A2	A2
IV-6	U	A1	A2	A2
IV-7	N	A2	A2	A2
IV-8	U	A1	A1/A2	A1/A2
IV-9	U	A1	-	A2
IV-10	U	A1/A2	A2	A2
IV-15	U	A1/A2	A1	A1/A2
IV-18	U	A1/A2	A1	A1

Note

a
- N — normal man
- MR — man with moderate mental retardation
- C — obligate carrier woman
- U — woman of unknown carrier status

Figure 3-3.

The multipoint likelihood function for the location of MRX1 on the background genetic map. The arrows indicate the points on the map at which location scores were initially calculated for each resample. The symbols within the figure show the range of resampled MRX1 locations under Protocol I (circles) and Protocol II (squares).



interval extended into the intervals DXS7-DXS14 and DXS14-DXYS1.

On resampling with both Protocol I and Protocol II and performing multipoint linkage analyses it was initially found that the peak location score occurred at DXS7 in some of the resamples, suggesting that MRX1 could be located at that point. Only two individuals in the pedigree (III-4 and III-21) had recombination between MRX1 and DXS7. Inclusion of either individual at least once in each resample ensured that the discontinuity observed at DXS7 in the original data was observed in each resample. This 'non-conservative' approach to resampling is discussed by Wilson and La Scala (1989; and the references therein).

For the 49 resamples under Protocol I, 46 had maxima at 0cM relative to DXS14. The three resamples with maxima away from DXS14 located MRX1 at -1.6, -3.3 and -5.1cM from DXS14. The range of peak location scores was 11.25-27.46. For protocol II, 41 of the 49 resamples located MRX1 at 0cM relative to DXS14. The remaining 8 resamples located MRX1 at -1.6cM (5 resamples), -3.3cM (1 resample), -6.9cM (1 resample), and -8.7cM (1 resample) from DXS14. The range of peak location scores for 47 of the resamples was 11.81 - 24.09; the remaining two location scores were 6.93 and 5.77 (both resamples placed MRX1 at 0cM). The resamples under both protocols which placed MRX1 at DXS14 were examined in detail.

In no resample was MRX1 placed in the interval DXS14-DXYS1. The resamples that placed MRX1 away from DXS14 were not re-examined in detail.

Under Protocol I the approximate 98% confidence interval for MRX1 location was 0 to -5cM from DXS14 i.e. extending distally from DXS14 towards DXS7. Under Protocol II the approximate 98% confidence interval was 0 to -9cM from DXS14.

Risk estimation

There were 13 women in this pedigree who were of unknown carrier status (Table 3-C). On the basis of pedigree data alone the carrier risks for these women ranged between 0.17 and 0.50. The risk intervals estimated with pedigree information and DXS14 genotypes were calculated for five values of the recombination fraction in the range 0.0-0.20 (the approximate 90% confidence interval for the recombination fraction between DXS14 and MRX1). Two-point risk analysis modified the carrier risk significantly for 11 of the women but the risk intervals were wide for a number of them. The two women whose carrier risks were not modified by two-point risk analyses (II-6 and IV-15) had mothers who were uninformative for DXS14.

For multipoint risk analyses, the carrier risk intervals were calculated for five MRX1 locations within the resampled approximate 98% confidence interval for gene location. As the

Table 3-C.

Carrier risks for women of unknown carrier status.

Individual	Pedigree alone[a]	Two-point risk analysis[b]	Multipoint risk analysis[c]	Multipoint risk analysis[d]
II-6	0.33	0.35-0.44	0.05-0.13	0.01-0.14
III-3	0.33	0.00-0.11	<0.01	<0.01
III-6	0.50	0.00-0.20	<0.01	<0.01
III-15	0.50	0.00-0.20	<0.01	<0.01
III-16	0.50	0.00-0.20	<0.01	0.00-0.32
IV-4	0.50	0.80-1.00	0.92-1.00	0.85-1.00
IV-5	0.50	0.80-1.00	0.92-1.00	0.85-1.00
IV-6	0.50	0.00-0.20	0.00-0.08	0.00-0.15
IV-8	0.17	0.00-0.06	<0.01	<0.01
IV-9	0.25	0.00-0.10	<0.01	<0.01
IV-10	0.25	0.00-0.10	<0.01	<0.01
IV-15	0.50	0.50	0.35-0.38	0.33-0.39
IV-18	0.25	0.00-0.04	<0.01	0.00-0.10

a Risks estimated using pedigree data alone;
b Risk intervals estimated using DXS14 genotypes for 5 values of the recombination fraction in the range 0.00 - 0.20;
c Risk intervals estimated using DXS7, DSX14, and DXYS1 genotypes with MRX1 located at 5 points within the resampled approximate 98% confidence interval for MRX1 location;
d Risk intervals estimated using DXS7, DSX14, and DXYS1 genotypes with MRX1 located at 5 points within the "one-LOD-unit-down" confidence interval.

confidence interval was wider with Protocol II than with Protocol I, the wider confidence interval was used for risk analysis.

The carrier risk intervals for all the women were narrowed. For eight of the women the carrier risk was less than 0.01. Although the mother of II-6 was uninformative for DXS14, she was informative for DXS7 and DXYS1 and multipoint risk analysis reduced the carrier risk significantly for II-6. Despite the mother of IV-15 being uninformative for both DXS14 and DXYS1, the carrier risk for IV-15 was reduced with the carrier risk interval being narrow. The mother of the sisters IV-4, IV-5 and IV-6 was uninformative for DXS7 and DXYS1. Despite this, the approximate confidence interval for MRX1 location indicated much narrower risk intervals for the sisters than had been estimated on the basis of two-point linkage analysis.

Risk intervals were also calculated using the "one-LOD-unit-down" confidence interval for MRX1 location. The "one-LOD-unit-down" confidence interval extended over two intervals, DXS7-DXS14 and DXS14-DXYS1. None of the women were recombinants in the interval DXS7-DXS14. Two women, III-16 and IV-18, were recombinants in the interval DXS14-DXYS1. As expected, the use of the "one-LOD-unit-down" confidence interval for multipoint risk analysis significantly widened the risk intervals for these two women.

DISCUSSION

MRX1 location

The locus causing the non-specific form of X-linked mental retardation in this family was located at Xp11. There was no evidence of linkage to F9 or DXS52 which flank FRAXA. Thus, MRX1 is not allelic to FRAXA.

This was the first clear evidence that non-specific X-linked mental retardation is genetically heterogenous (Suthers et al. 1988a [Appendix D]). Choo et al. (1984) had earlier described two families with non-specific X-linked mental retardation and normal karyotypes where the disease loci were not linked to F9. But in view of the proposed linkage heterogeneity between FRAXA and F9 (Brown et al. 1985, 1986, 1987b, 1988), the lack of linkage between these loci and F9 did not exclude the possibility that the loci were allelic to FRAXA.

The genetic heterogeneity of non-specific X-linked mental retardation has now been confirmed by others. The loci causing non-specific X-linked mental retardation in two other families have been mapped to Xp22.2-p22.3 and Xq12-q13, with suggestive evidence of a third locus at Xp21.2 (the peak LOD score was 1.59)(Arveiler et al. 1988b).

Confidence interval for MRX1 location

The small number of meioses in this pedigree places a limit on the precision with which MRX1 can be localized. It is not possible to pool data from a number of families as there is now clear evidence that non-specific X-linked mental retardation is genetically heterogeneous. The precision of the localization was greater with multipoint linkage analysis than with two-point linkage analysis, but there was no method for estimating a confidence interval of known significance for gene location.

A resampling or "bootstrap" method was used for estimating an approximate confidence interval for gene location on a known genetic map using genotype data from a single pedigree. In using this resampling method to determine the approximate 98% confidence interval for MRX1 location, MRX1 was consistently located in the interval DXS7-DXS14. In contrast the "one-LOD-unit-down" method suggested that the confidence interval was -17 to +5cM relative to DXS14 i.e. extending over the intervals DXS7-DXS14 and DXS14-DXYS1.

For a given background genetic map the resampled estimate of an approximate confidence interval for gene location appeared robust i.e. the same estimate was obtained using different

resampling protocols. However, two cautions should be noted[1]. First, the resampling protocol must be adapted for each situation being analyzed. Moreover, the appropriate number of resamples (B) must be determined for each situation. In analyzing this pedigree the results were not qualitatively altered by increasing the number of resamples from 19 to 49 (Chapter 2). This contrasts with the situation examined by Wilson and La Scala (1989) where there was a change in the bounds of the confidence interval as the number of resamples increased from 19 to 49.

Secondly, sophisticated data analysis cannot replace appropriate data collection. Genotypic data from other affected males in the pedigree or for other markers in the interval DXS7-DXS14 would alter the confidence interval for MRX1 location. The highly polymorphic locus DXS255 (Kidd et al. 1989) lies in the interval DXS7-DXS14 (Keats et al. 1989) but unfortunately the pedigree was uninformative for this marker.

Risk estimates

The ability to define an approximate confidence interval for gene location has implications for genetic counselling. In the case of a common genetic disorder, the confidence

---

1 A third caution is that changes in the background genetic map may radically alter the size of the estimated confidence interval. This issue is discussed by Suthers & Wilson (1990; [Appendix D]).

interval is usually narrow, and risk estimates based on a single disease gene location provide sufficient information for consultands. However, many rare conditions have been mapped in only a few pedigrees. In this situation the confidence interval for gene location will usually be wide, and the risk estimates provided to consultands must reflect the uncertainty of the gene location (Lange 1986). It is now possible to define the uncertainty of multipoint risk estimates by using the resampling method to estimate a confidence interval for gene location.

For each consultand the estimate of carrier risk is a likelihood function which varies with gene location. This function need not be monotonic, as Krawczak (1987) demonstrated for two-point data. He showed that it is not sufficient to calculate two-point risk estimates for just the two confidence limits for the recombination fraction since there may be intermediary higher or lower risk estimates. In the case of multipoint risk analysis, the approximate confidence interval for gene location indicates the range of gene location values for which risks should be determined.

The combination of multipoint risk analysis and an appropriate confidence interval for gene location can change significantly the estimated risk intervals provided to a consultand. Incorporating interference in multipoint risk analysis may narrow the risk interval even further by

reducing the possibility of double recombination. In the pedigree described, none of the women of unknown carrier status were definite recombinants between the polymorphic loci which flanked MRX1, and the inclusion of interference would have had little effect on the width of the carrier risk intervals.

Conclusion

The arbitrary structure of human pedigrees and the complexity of multipoint linkage analysis usually make it impossible to provide numerical estimates of the reliability of the results obtained. The resampling method is a powerful non-analytical approach to estimating the reliability of linkage results, and can be applied in the study of very rare genetic conditions where there is a limited amount of data. The application of this method to a linkage study of a single pedigree has shown that non-specific X-linked mental retardation is genetically heterogeneous, and has had immediate clinical benefit in counselling women in the family.

Chapter 4.

PHYSICAL MAPPING OF NEW DNA PROBES NEAR THE FRAGILE X (FRAXA) WITH A PANEL OF CELL LINES.

SUMMARY

A panel of 14 somatic cell hybrid lines, lymphoblastoid cell lines, and peripheral lymphocytes with X chromosome translocation or deletion breakpoints near FRAXA is described. The locations of the breakpoints were defined with 14 established probes located between DXS100 and DXS52.

Seven of the cell lines had breakpoints between the loci DXS369 and DXS304 which flank FRAXA at a distances of 3-5 centimorgans. The gene responsible for Hunter syndrome (IDS deficiency) was located at the breakpoint of one of these cell lines, thus localizing IDS to the same interval as FRAXA.

The panel of cell lines was used to localize 18 new DNA probes in this region. Eight of the probes, VK16, VK18, VK21, VK23, VK24, VK37, VK47, and pc2S15 (an IDS cDNA probe), detected loci near FRAXA and it was possible to order both the X chromosome breakpoints and the probes in relation to FRAXA. The order of probes and loci near FRAXA is :
cen-DXS369,VK24-VK47-VK23-VK16,FRAXA-VK21-VK18-IDS-VK37-DXS304-qter.

INTRODUCTION

The locus responsible for the fragile X syndrome (FRAXA) is located at or very near the characteristic fragile site at Xq27.3. The development of a precise genetic map in this region of the X chromosome has been hampered by a lack of closely linked polymorphic loci. Until recently the closest polymorphic loci to FRAXA, DXS105 (cX55.7), DXS98 (4D-8), and DXS52 (St14-1), lay more than 10 centiMorgans (cM) from FRAXA (Mandel et al. 1989).

In 1989 two polymorphic loci within 5 cM of FRAXA were reported, DXS369 (RN1), and DXS304 (U6.2)(Hupkes et al. 1989; Vincent et al. 1989). The established order of probes and genes near FRAXA was:

cen-F9-DXS105-DXS98-DXS369-FRAXA-DXS304-DXS52-qter

(Mandel et al. 1989). A key factor in the localization of DXS369 and DXS304 was the mapping of each locus using somatic cell hybrids before the search for restriction fragment length polymorphisms (RFLPs) was undertaken. A panel of cell lines with well-defined X chromosome breakpoints near FRAXA would allow for the rapid identification of more probes close to FRAXA and the further development of precise genetic and physical maps around FRAXA.

In this chapter a panel of somatic cell hybrid lines, lymphoblastoid cell lines, and peripheral lymphocytes that

contain human X chromosomes with deletion or translocation breakpoints near FRAXA is presented. This panel of cell lines was used to physically localize a series of new DNA probes near FRAXA. Mapping of these probes in relation to FRAXA by *in situ* hybridization defined the locations of the X chromosome breakpoints and the other probes relative to FRAXA. Eight of the new probes and the breakpoints in seven of the cell lines were located close to FRAXA in the interval DXS369-DXS304. The gene responsible for Hunter syndrome (iduronate sulfatase [IDS] deficiency; mucopolysaccharidosis type II) was mapped to immediately distal to FRAXA.

METHODS

The cell type, cytogenetic description, source, and references for each cell line are summarized in Table 2-A (Chapter 2).

The loci used to map the breakpoints in the cell lines are listed in Table 4-A. The positions of these loci on the X chromosome, the coresponding DNA probes, source, and references are summarized in Table 2-C (Chapter 2).

The isolation of the series of 17 new DNA probes (all with the prefix "VK") has been described (Hyland et al. 1989; Chapter 2). The locus names for these probes are listed in Table 2-B (Chapter 2).

The probe pc2S15 was a 1.5 kb IDS cDNA clone recently isolated by PJ Wilson (Adelaide, S.A.). An oligonucleotide had been synthesized on the basis of the peptide sequence of the NH_2-terminal of purified iduronate sulfatase enzyme. This oligonucleotide probe was used to isolate a genomic DNA fragment containing the 5′-end of the IDS gene. The genomic fragment was used to isolate a series of overlapping cDNA clones from a human endothelial cDNA library. pc2S15 was a cDNA clone which contained over 90% of the coding sequence of IDS (Wilson et al., submitted). The studies of DNA from boys with Hunter syndrome using pc2S15 were performed by PJ Wilson

Note to Table 4-A:

DNA probes were used to determine whether the corresponding loci were present (+) or absent (-) in the specified cell lines. This composite Table lists results using both the established probes and the new probes. The established probes (and corresponding loci) and cloned genes are listed in order down the X chromosome. The VK probes, pc2S15, and CDR-9 are listed in the order defined by the locations of adjacent breakpoints on the X chromosome. This table includes both data published on individual cell lines (references in Table 2-A) and previously unpublished data. The cell line "B17" refers to cell line 908K1B17. The breakpoint in CY34 is located at IDS (shown as +/-).

Table 4-A.
Location of breakpoints in cell lines determined by Xq probes.

	CY2	CY3	LL556	GM08121	LC12K15	2384-A2	B17	TC4.8	PeCH-N	APC-5	04-1	03-1	CY34	Y162.Aza
pX45d (DXS100)	+	-					-		+				+	
HPRT	-	+				+				+				+
52A (DXS51)			+				-			+				
VK10 (DXS290)	-	+	+		+	+	-	-						
VK17 (DXS294)	-	+	+		+	+	-	-	+				+	+
VK29 (DXS300)	-	+	+		+	+	-	-	+					
VK34 (DXS301)	-	+	+		+	+	-	-	+				+	
VK41 (DXS310)	-	+	+		+	+	-	-	+				+	
cX38.1 (DXS102)			-				-							
F9	-	+	-	+	+	+	-	-		+				+
VK11 (DXS291)	-	+	-	+	+	+		-						
cX55-7 (DXS105)	-	+	-	-	+	+	-	-	+	+			+	+
CDR-9 (CDR)			-	-	+	+	-	-					+	+
4D-8 (DXS98)	-	+	-	-	-	+	-	-	+	+	+	+	+	+
VK7 (DXS288)	-	+	-		-	+	-	-	+				+	+
VK14 (DXS292)	-	+	-		-	-	+	-	+				+	
RN1 (DXS369)	-	+	-		-	-	+	+	+	+	+	+	+	+
VK24 (DXS298)	-	+	-		-	-	+	+	+				+	
VK47 (DXS308)	-	+	+		-	-	+	+	+				+	
VK23 (DXS297)	-	+	+		-	-	+	+	-	+	+	+	+	+
VK16 (DXS293)	-	+	+		-	-	+	+	-	-	+	+	+	+
VK21 (DXS296)	-	+	+		-	-		+	-	-	-	+	+	+
VK18 (DXS295)	-	+	+		-	-	+	+	-	-	-	+	+	+
pc2S15 (IDS)	-	+	+		-	-	+	+	-	-	-	-	+/-	+
VK37 (DXS302)	-	+			-	-	+	+	-		+	+	-	
U6.2 (DXS304)	-	+	+		+	-	+	+	-	-	+	+	-	+
1A1 (DXS374)	-	+	+		+	-	+	+	-	-	+	+	-	+
VK25 (DXS299)	-	+			+				-				-	+
F8C				-	+	-	+	+		-				-
G6PD				-	+	-		+						-
MN12 (DXS33)					-		-	+						
St14-1 (DXS52)	-	+	+	-	+	-	+	+	-	-			-	-
VK9 (DXS289)	-	+	+		+	-	+	+	-				-	-

or PV Nelson (Adelaide, S.A.). Studies of cell line DNA using pc2S15 and of DNA from boys with Hunter syndrome using other probes were performed by the candidate.

The DNA methods are described in Chapter 2. For *in situ* hybridization studies (performed by E Baker and H Eyre, Adelaide S.A.), ^{3}H-labelled probes were hybridized (Simmers et al. 1988) to prephotographed metaphase chromosomes at concentrations of 0.01-0.1 ug/ml for 19-27 days. Probes containing repeated sequences were pre-reassociated with unlabelled total human DNA (Sealey et al. 1985). Silver grains that appeared to touch the X chromosome were scored.

RESULTS

Localization of the Hunter syndrome gene, *IDS*.

Cell line CY34 was derived from a girl with Hunter syndrome and a t(X;5) reciprocal translocation (Chapter 2), and it had been postulated that IDS was located at the breakpoint on the X chromosome. DNA probes that detect the established loci listed in Table 4-A were used to probe DNA from CY34. The presence or absence of the respective loci are listed in the table. CY34 retained loci proximal to (and including) DXS369, and had lost loci distal to DXS369, indicating that IDS was in the same interval as FRAXA.

When pc2S15 was used to probe DNA from CY34 it hybridized to only four of the eight fragments seen in DNA from normal controls (Fig. 4-1). This demonstrated that the translocation did disrupt IDS. *In situ* hybridization of pc2S15 to the chromosomes of a normal female confirmed the localization to Xq28 and established that there were no sequences homologous to pc2S15 elsewhere in the genome (Fig. 4-2).

When pc2S15 was hybridized to chromosomes expressing the fragile X, 79 grains could be scored relative to the fragile site. 45 grains lay between the fragile site and the telomere, 21 grains lay within a similar distance proximal to the fragile site, and 13 grains lay centrally over the fragile site. The difference between the number of proximal

Figure 4-1.
The pc2S15 probe detected sequences spanning the breakpoint in cell line CY34. pc2S15 was used to probe a Southern blot of HindIII-digested DNA from a normal female (lane 1), normal male (lane 2), and cell lines CY34 (lane 3), CY34A (lane 4), and A9 (lane 5). CY34 contained Xpter-q28 while CY34A was a subclone of CY34 having just Xq24-q28; both cell lines had a mouse (A9) background. In normal DNA pc2S15 detected fragments of 7.5 kb, 5.5 kb, 4.1 kb, 4.0 kb, 2.5 kb, 2.3 kb, 1.3 kb, and (very faintly) 0.76 kb. In CY34 and CY34A DNA the 5.5 kb, 2.5 kb, 1.3 kb and 0.76 kb fragments were not detected and a new 1.0 kb fragment was faintly visible. A very faint 14 kb fragment was visible in the A9 lane. pc2S15 did not detect polymorphic HindIII fragments in DNA samples from 16 normal X chromosomes. Approximate DNA fragment size indicators are shown on the left of the figure.

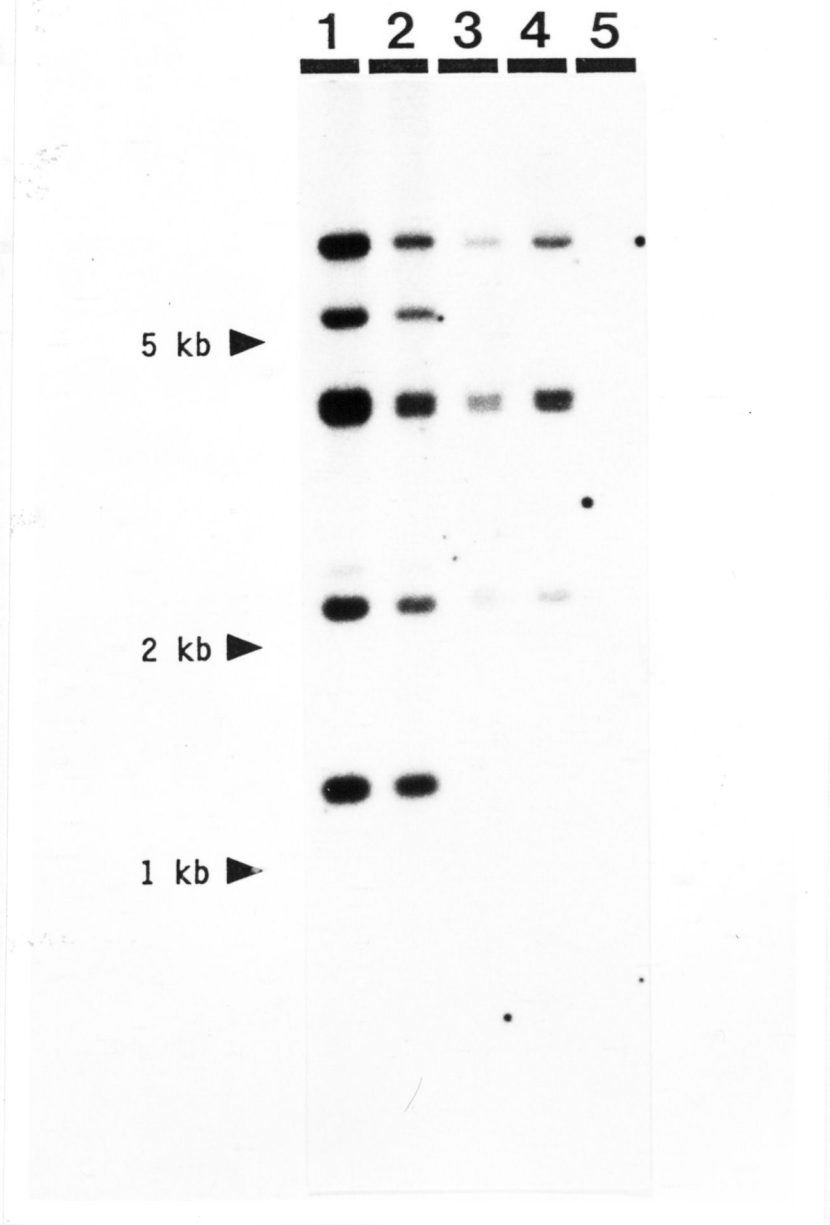


Figure 4-2(a).

In situ hybridization of pc2S15 to normal chromosomes. pc2S15 was ^{3}H-labelled and hybridized to metaphase chromosomes from a normal female. The silver grains which appeared to touch a chromosome were scored. 150 grains were scored from 30 metaphases. pc2S15 detected sequences at distal Xq only; sequences homologous to pc2S15 were not detected elsewhere on the X chromosome or on the autosomes. There was no significant hybridization to the Y chromosomes of two males (data not shown).

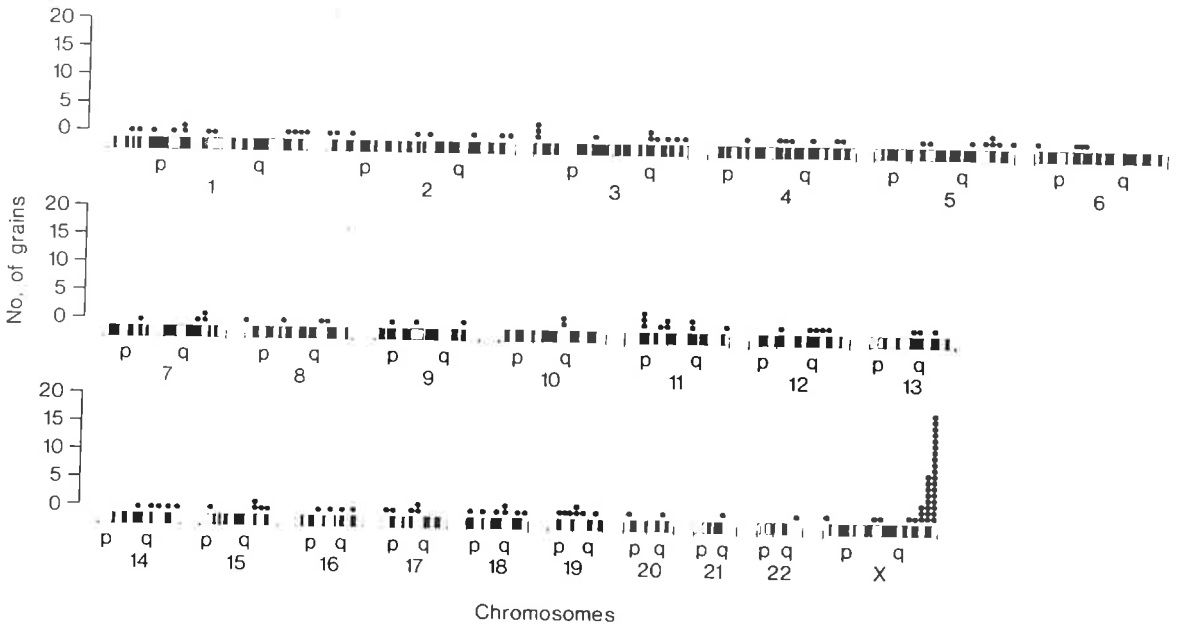
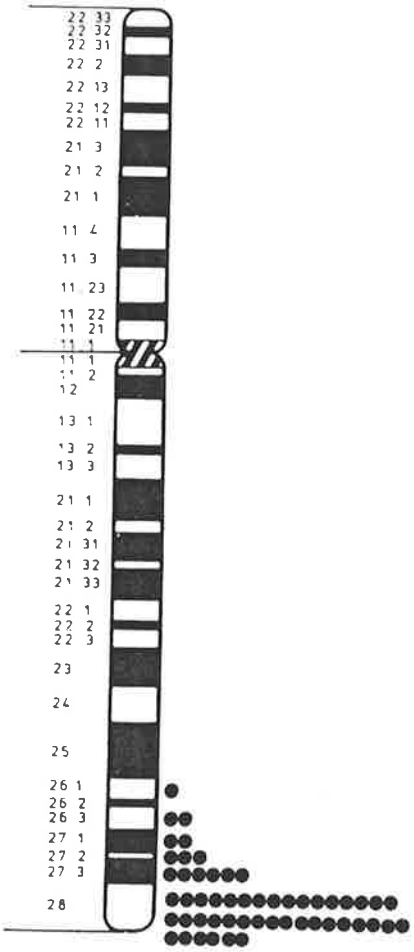


Figure 4-2(b).
An additional 40 metaphases with high resolution chromosome banding (600-1000 bands per metaphase) were scored and indicated that IDS is located at Xq28.



versus distal grains was significant ($X^2_1 = 8.73$; $p<0.005$). This indicated that IDS and the CY34 breakpoint were distal to FRAXA.

The genomic clone which contained only the 5'-end of IDS detected a 1.3 kb fragment in normal female DNA but did not detect any fragments in CY34 (data not shown), indicating that IDS was orientated with the 5'-end on the telomeric side of the CY34 breakpoint.

The pc2S15 probe was used to analyze genomic DNA from 23 unrelated British and Australian males who had complete absence of IDS activity on enzymatic assay (Hopwood 1989). Seven individuals had structural alterations of IDS; two (03-1 and 04-1) had deletions of the entire pc2S15 coding region while five had various partial deletions or re-arrangements (Wilson et al. 1990 [Appendix D]). The loci DXS304 and DXS369 were not deleted in any of these patients (data for 03-1 and 04-1 are listed in Table 4-A).

Mapping translocation and deletion breakpoints in cell lines

DNA of the other cell lines was probed to determine whether the established loci listed in Table 4-A were present. The probes used for each cell line were selected to delineate the breakpoints on the X chromosome. The presence or absence of the respective loci in each sample are listed in Table 4-A. Two of the cell lines, LL556, and LC12K15, had interstitial

deletions; the suffix 'p' or 'd' will be used to indicate the proximal or distal breakpoints in these cell lines.

Five cell lines, PeCH-N, LL556d, APC-5, CY34, and LC12K15d, had breakpoints between the loci which flank FRAXA, DXS369 and DXS304. In addition, 03-1 and 04-1 had interstitial deletions in the same interval. Using the established probes it was not possible to determine the order of the breakpoints either in relation to each other or to FRAXA. Three cell lines had breakpoints between DXS98 and DXS369, proximal to FRAXA.

There were no inconsistencies in these data to suggest complex rearrangements in the cell lines in the region Xq26-28. The human chromosomal component of the cell line APC-5 is known to be fragmented (Chapter 2) but the data in Table 4-A do not indicate an interstitial deletion or rearrangement at Xq26-28. Assuming that complex rearrangements were not present, the locations of translocation breakpoints and the locations and extent of interstitial deletions in the cell lines are summarized in Figure 4-3.

Mapping new DNA probes using the cell line panel

Seventeen VK probes were mapped using the panel of cell lines (Fig. 4-4). The presence or absence of the respective loci in each cell line is listed in Table 4-A and summarized in Fig. 4-3.

Figure 4-3.

The locations of cell line breakpoints, established DNA probes, cloned genes, and the VK probes at Xq26-28. The cell line translocation and deletion breakpoints are indicated to the left of the X chromosome. For each cell line the arrow indicates on which side of the breakpoint lies the X chromosome material retained in the cell line. In the case of cell lines with interstitial deletions the proximal (p) and distal (d) breakpoints are shown. The established probes and cloned genes are shown to the right of the chromosome in order down the X chromosome. The locations of the VK probes, pc2S15 (IDS), and CDR-9 are shown on the far right. The locations of all the probes and the cell line breakpoints in relation to chromosome banding is approximate. The fragile X is represented as a hatched region at distal Xq27; the locations of the APC-5 and 04-1p breakpoints and of the probe VK16 relative to the fragile X are unknown.

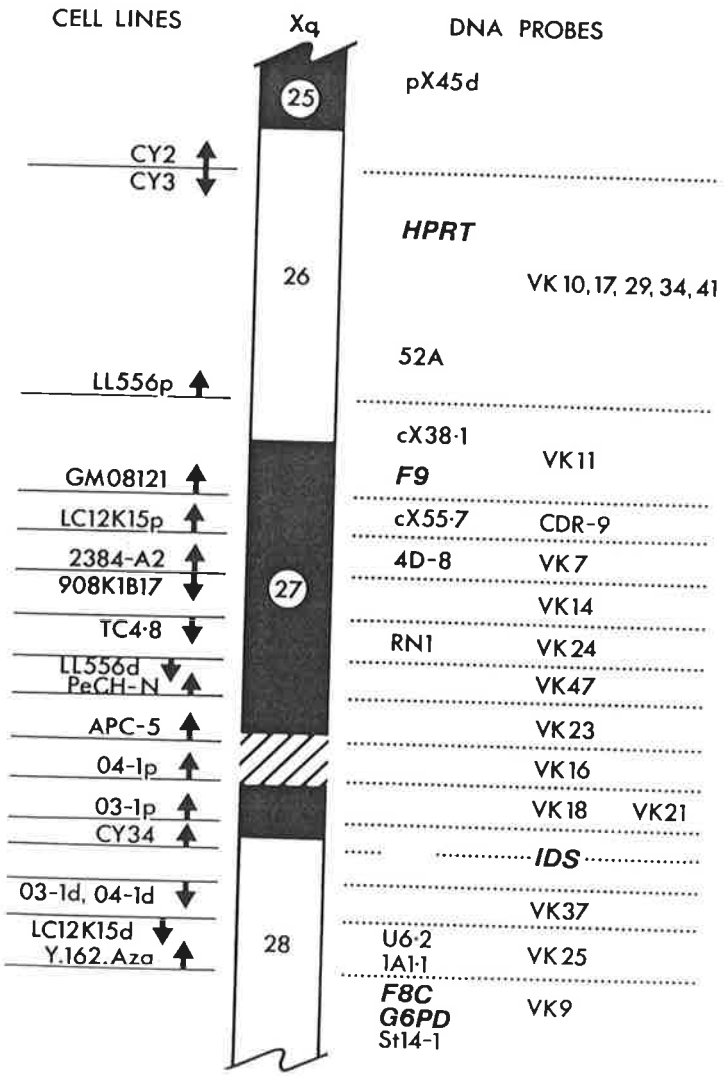


Figure 4-4.

Physical mapping of VK clones using various cell lines. Each VK probe was radiolabelled, pre-reassociated, and hybridized to HindIII-digested DNA extracted from cell lines. Examples of the autoradiographs obtained during these experiments are presented. Two exposures were obtained of each autoradiograph. Discrete bands were usually evident in the lanes containing total human DNA within 1-2 days; by the time discrete bands were evident in the lanes containing DNA from somatic cell hybrids (3-10 days), lane background due to repeated sequences had obliterated the signal in the lanes of total human DNA.

(a) VK9 hybridized to DNA from a normal male and from various cell lines. VK9 hybridized to DNA from PeCH-A, 908K1B17, LC12K15, LL556 (obscured in this exposure), and TC4.8. (PeCH-A was the reciprocal transloaction product of PeCH-N, and contained Xq27-qter. PeCH-A was used infrequently, and the results are not included in Table 4-A). VK9 did not hybridize to DNA from cell lines 2384 or Y162.Aza. These results indicated that VK9 detected a locus distal to the Y162.Aza breakpoint.

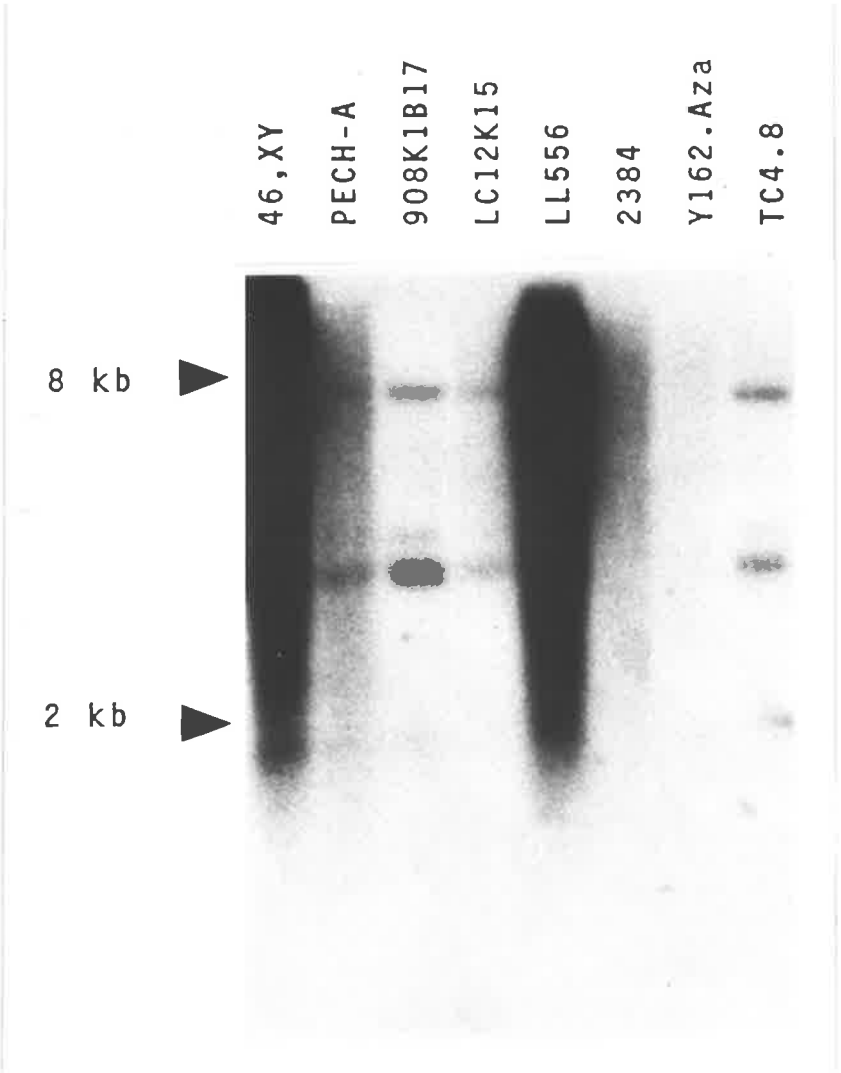


Figure 4-4(b)
VK16 hybridized to DNA from a normal male and from cell lines CY3 and CY34; VK16 did not hybridize to DNA from cell lines CY2 and PeCH-N. This indicated that VK16 detected a locus between the breakpoints in the cell lines PeCH-N and CY34.

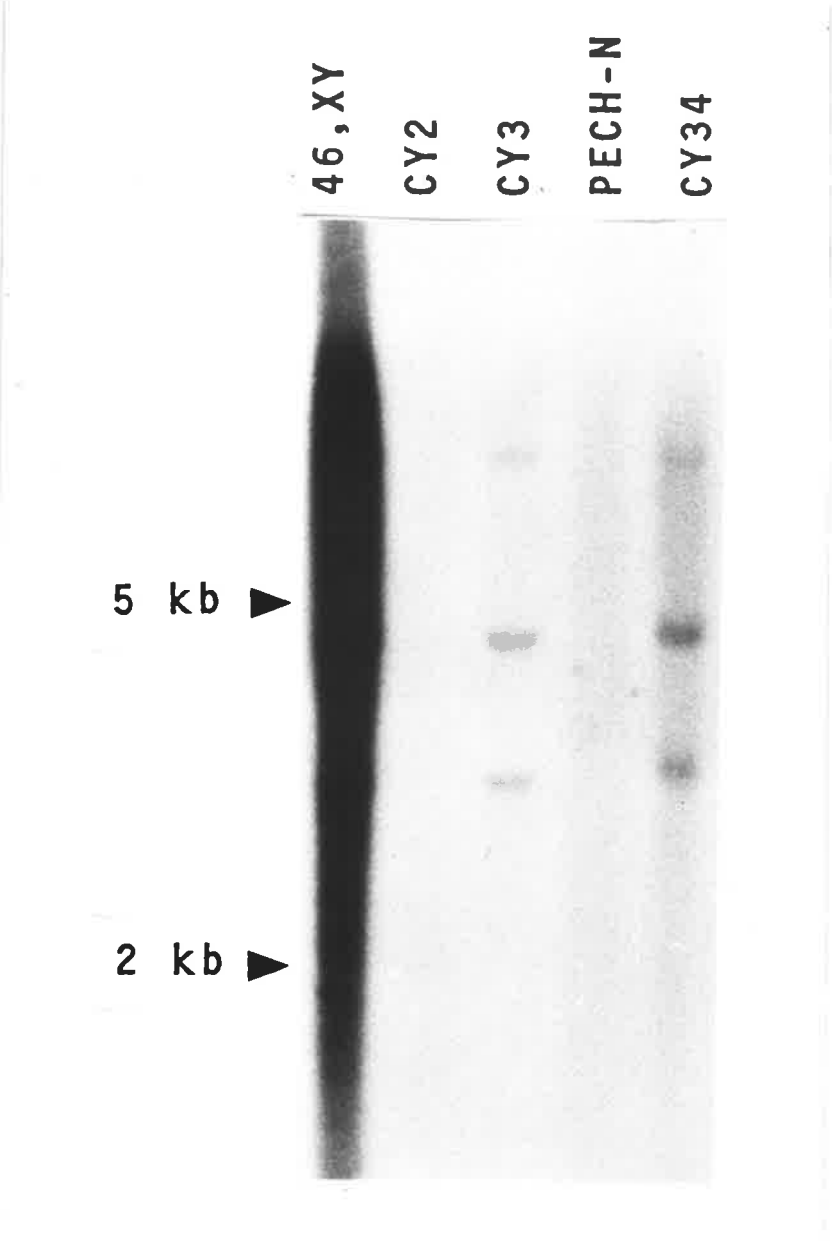


Figure 4-4(c).
VK18 hybridized to DNA from normal females and from cell lines CY3 and CY34; VK18 did not hybridize to DNA from cell lines CY2 and PeCH-N. This indicated that VK18 detected a locus between the breakpoints in the cell lines PeCH-N and CY34.

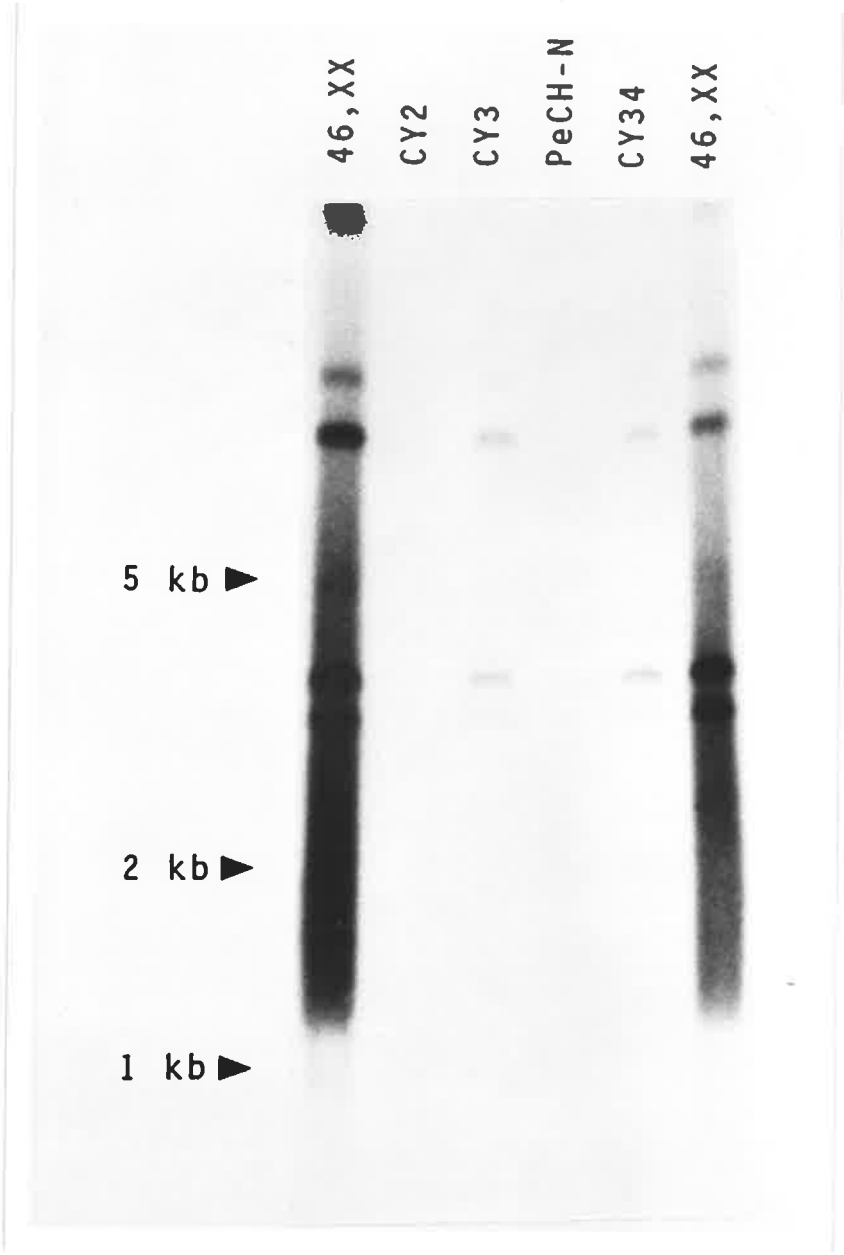


Figure 4-4(d).
VK23 hybridized to DNA from a normal male and from cell lines CY3 and CY34; VK23 did not hybridize to DNA from cell lines CY2 and PeCH-N. This indicated that VK23 detected a locus between the breakpoints in the cell lines PeCH-N and CY34.

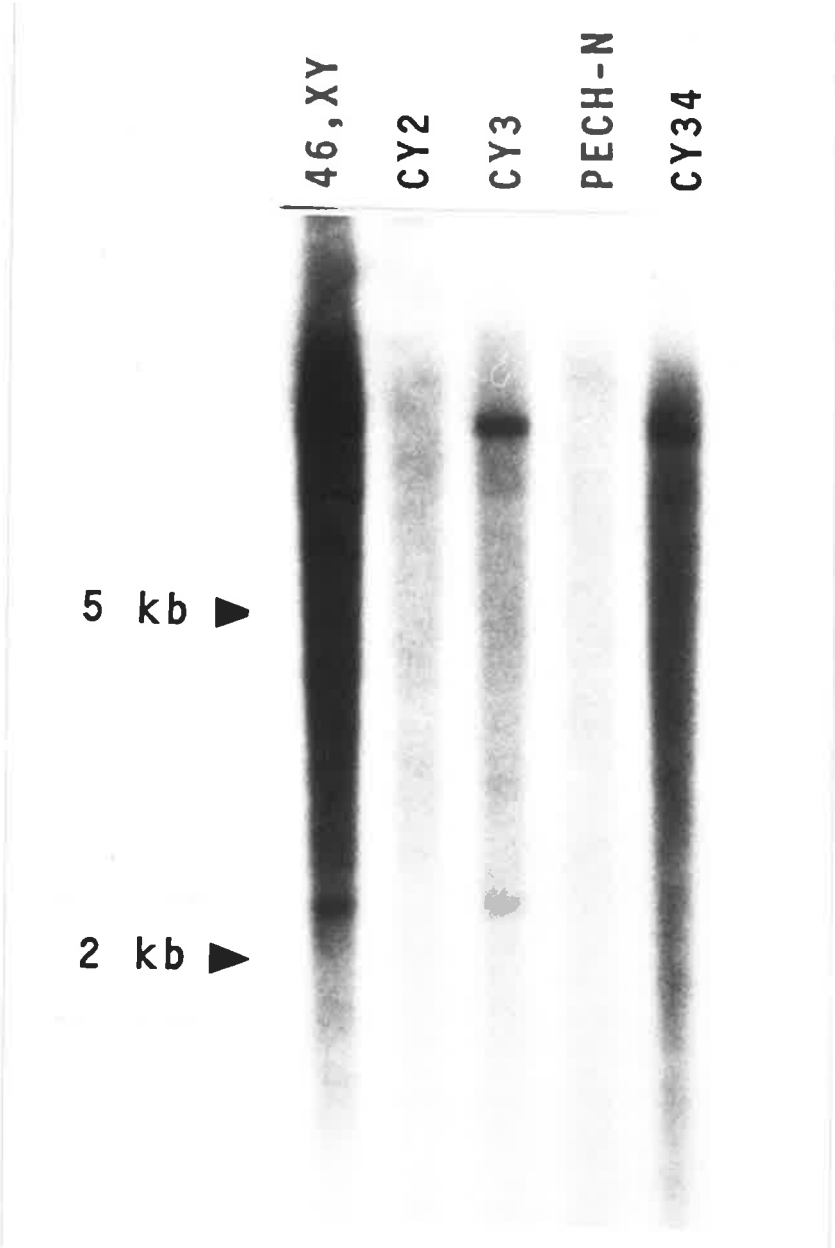
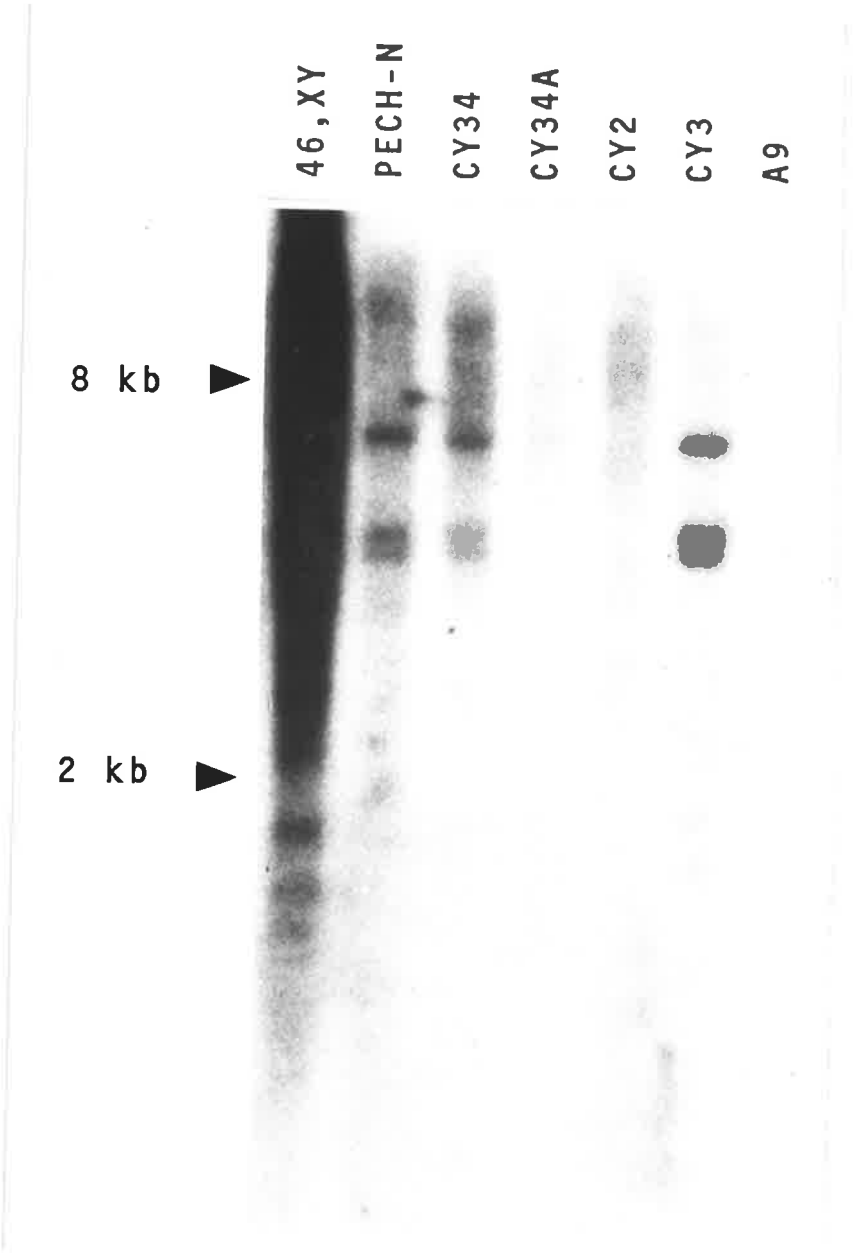


Figure 4-4(e).
VK41 hybridized to DNA from a normal male and from cell lines PeCH-N, CY34, and CY3; VK41 did not hybridize to DNA from cell lines CY34A, CY2 and A9. This indicated that VK41 detected a locus included in an interstitial deletion in CY34A at Xq26.



Six of the VK probes were located close to FRAXA between the loci DXS369 and DXS304. One probe was located in the same interval as DXS369. Eight probes were located proximal to DXS369, one was in the same interval as DXS304 and DXS374, and one mapped distal to DXS374.

The VK probes indicated the order of some of the breakpoints that could not be resolved using the established probes. The results with VK probes that detected loci close to FRAXA were of particular interest. The probes VK16 (DXS293), VK18 (DXS295), VK21 (DXS296), VK23 (DXS297), VK37 (DXS302), and VK47 (DXS308) were located between DXS369 and DXS304 and resolved the locations of the breakpoints in that interval. The order of breakpoints between DXS369 and DXS304 down the chromosome was LL556d/PeCH-N/APC-5/04-1p/03-1p/CY34/03-1d,04-1d/LC12K15d. The probe VK14 (DXS292) lay between DXS98 and DXS369 and separated the TC4.8 breakpoint from the 908K1B17 and 2384-A2 breakpoints.

The CY34 breakpoint had been placed distal to FRAXA by *in situ* hybridization studies of pc2S15 (detailed above). The locations of VK21 and VK23 in relation to FRAXA were determined in the same manner. A 1.3 kb HindIII-SalI fragment of VK21 (VK21A) which contained few repeated sequences was hybridized to metaphase chromosomes expressing the fragile X. Of the 139 silver grains that touched the X chromosome and

could be scored relative to the fragile X, 70 grains were distal and lay between the fragile site and the telomere, 43 grains lay within a similar distance proximal to the fragile site, and 26 grains were located centrally over the chromosome gap at the fragile site. The difference between the number of proximal versus distal grains was significant (X^2_1=6.45; p<0.02). This indicated that VK21 was distal to FRAXA, and that the 03-1p breakpoint was also distal to FRAXA.

A 2.6 kb EcoRI-PstI fragment of VK23 (VK23B1) containing few repeated sequences was isolated. VK23B1 was also hybridized to chromosomes expressing the fragile X. Of the 42 silver grains that could be scored relative to the fragile site, 8 grains lay between the fragile site and the telomere, 31 grains lay within a similar distance proximal to the fragile site, and 3 grains were located centrally over the chromosome gap at the fragile site. The difference in the number of proximal versus distal grains was significant (X^2_1= 13.56; p<0.0005) indicating that VK23 was proximal to FRAXA. This in turn indicated that VK47 and the PeCH-N and LL556d breakpoints were proximal to FRAXA.

The location of the probe VK11 (DXS291) in relation to the breakpoint in cell line GM08121 was determined by *in situ* hybridization. There was no significant difference in the number of grains on the normal versus deleted X chromosomes

(normal/deleted chromosomes=24/23 grains; X^2_1= 0.02, p>0.10) indicating that VK11 was proximal to the breakpoint.

The cell line CY34A was a subclone of CY34 that contained just a fragment of the original derived X chromosome extending from Xq24 to Xq28 (Chapter 2). The probes VK29 (DXS300), VK34 (DXS301), and VK41 (DXS310) detected an interstitial deletion in CY34A. These probes detected the corresponding loci in cell line CY34 (Table 4-A) but did not detect the loci in CY34A (data not shown). On the basis of results from other cell lines the three probes could be localized to Xq26. This indicated that CY34A had an interstitial deletion at Xq26. Two other probes in this region, VK10 (DXS290) and VK17 (DXS294), detected loci in both CY34 (Table 4-A) and CY34A (data not shown).

The gene probe for a cerebellar-degeneration-related protein, CDR-9, has been localized to the interval between the GM08121 and 2384-A2 breakpoints (Hirst et al. 1990). CDR-9 detected sequences in the cell line LC12K15 (Table 4-A) thus localizing CDR proximal to the LC12K15p breakpoint in the same interval as the locus DXS105.

Two of the VK probes detected conserved sequences in human, mouse, and hamster DNA. VK21 detected a single conserved fragment in HindIII-digested DNA of the mouse (1.9 kb) and hamster (1.8 kb)(data not shown). VK25 also detected a single

fragment in mouse and hamster DNA that was easily distinguished from the human fragments.

DISCUSSION

Localization of new DNA probes near *FRAXA*

Somatic cell hybrids containing human X chromosomes with translocation or deletion breakpoints have been used to physically map DNA probes in various regions of the long arm of the human X chromosome (Wieacker et al. 1984; Oberle et al. 1986; Hofker et al. 1987; Cremers et al. 1988). Few cell lines have been described with breakpoints close enough to FRAXA to be useful in localizing new probes near this locus. In this chapter a number of cell lines are presented which have precisely defined breakpoints close to FRAXA, and a series of DNA probes are mapped close to FRAXA (Fig. 4-3).

The availability of these cell lines makes it feasible to rapidly localize clones from a DNA library and to identify those clones derived from regions physically close to FRAXA. A total of 13 breakpoints in 10 cell lines are now defined between DXS105 and DXS304. Seven of these cell lines had breakpoints between DXS369 and DXS304. The cell line with the closest breakpoint known to be proximal to FRAXA was PeCH-N; the closest breakpoint known to be distal to FRAXA was 03-1p. The APC-5 and 04-1p breakpoints were within the interval defined by the PeCH-N and 03-1p breakpoints, but the locations of the APC-5 and 04-1p breakpoints in relation to FRAXA are unknown.

Two intervals between DXS369 and DXS304 contained more than one probe, DXS369 with VK24 (DXS298), and VK21 with VK18. Probes that detect RFLPs can be ordered by genetic linkage studies, but if the probes are very close together it becomes increasingly unlikely that recombination between the loci will be observed. It was possible to order the probes near FRAXA by pulsed field gel electrophoresis. VK21, VK18, IDS, and DXS304 have been linked together on a large scale restriction map with VK18 placed distal to VK21 (Hyland et al., manuscript submitted). Since DXS304 lies 3 cM distal to FRAXA (Mandel et al. 1989) this map encompasses five probes (including VK37) and five breakpoints in a region immediately distal to FRAXA.

It may be possible to generate a similar map of the region proximal to FRAXA. DXS369 lies 5 cM proximal to FRAXA (Oostra et al. 1990) and the region between DXS369 and VK21 now encompasses five probes (and possibly a sixth, VK24), four breakpoints, and FRAXA. The development of the large scale restriction map around FRAXA would be further enhanced by the use of probes from linking or jumping libraries that had been localized with the cell panel.

The order of probes and genes near FRAXA is now cen-DXS369,VK24-VK47-VK23-VK16,FRAXA-VK21-VK18-IDS-VK37-DXS304-qter. The fact that these new VK probes could be localized and ordered near FRAXA without genetic linkage studies

demonstrates the value of this panel of cell lines in further investigation of FRAXA. If RFLPs are detected by the VK probes they will be valuable in developing the fine scale genetic linkage map around FRAXA. A polymorphism detected by VK16 would be particularly useful in both analyzing the unusual segregation of the fragile X mutation (Sherman et al. 1985) and in providing genetic advice in affected pedigrees.

Localization of the Hunter syndrome gene

Linkage studies had suggested that the Hunter syndrome gene, IDS, was located at Xq26-q28 (Roberts et al. 1989). However, the LOD scores were relatively low, and the position of IDS in relation to the polymorphic loci at Xq27 was not known. On the basis of the physical mapping presented in this chapter, IDS is located between VK21 and DXS304. The order of loci is cen-FRAXA-VK21-IDS-DXS304-qter (Fig. 4-3), and DXS304 is 3 cM distal to FRAXA (Vincent et al. 1989b). Therefore, VK21 and DXS304 flank IDS at distances of less than 3 cM.

The boys 03-1 and 04-1 had extremely severe features of Hunter syndrome (Wilson et al. 1990 [Appendix D]; Chapter 2). The boy 04-1 had a deletion that included IDS, VK18, and VK21, and extended towards FRAXA (Table 4-A). VK21 detects conserved sequences in human, mouse, and hamster DNA. The conservation of X chromosome coding sequences (Ohno 1969) suggests that VK21 could detect an X-linked gene or pseudogene. It is conceivable that the deletion in 04-1

encompassed genes other than IDS. The boy 04-1 did not have any congenital malformations, although the severity of his Hunter syndrome phenotype made it difficult to exclude either mild congenital facial dysmorphism or congenital mental retardation. It will be important to carefully correlate the extent of large deletions around IDS with the patients' phenotypes to define any contiguous gene syndromes (Schmickel 1986).

The general availability of lymphoblastoid or fibroblast lines from such patients would be a valuable resource for mapping other new probes near IDS and FRAXA. At present there is no evidence to suggest that FRAXA is itself a deletion (Sutherland et al. 1985; Laird et al. 1987) and DNA from patients with Hunter syndrome is currently the best potential source of interstitial deletions near FRAXA.

Conclusion

The utility of these cell lines is not limited to studies of the fragile X syndrome. An increasing number of disorders are being localized to Xq26-28 (Mandel et al. 1989). The rapid mapping of DNA probes within a small region of the human genome with a panel of cell lines opens up the possibility of isolating yeast artificial chromosomes (YACs) that overlap in the region, developing a large scale restriction map, localizing conserved sequences or cDNA probes, and ultimately isolating the gene of interest.

Chapter 5.

LINKAGE ANALYSIS OF NEW RFLPS AT Xq27-Q28 IN NORMAL FAMILIES.

SUMMARY

The order of loci near FRAXA has been defined on the basis of physical mapping studies to be cen-DXS369,VK24-VK47-VK23-VK16,FRAXA-VK21-VK18-IDS-VK37-DXS304-qter. Restriction fragment length polymorphisms (RFLPs) have been identified at DXS369 and DXS304. These two loci lie within 5 cM of FRAXA.

The probe VK21 detected TaqI, MspI, and BclI RFLPs at DXS296 with heterozygote frequencies of 0.23, 0.31, and 0.31 respectively. The probe VK23 detected HindIII and XmnI RFLPs at DXS297 with heterozygote frequencies of 0.34 and 0.49 respectively. The IDS cDNA probe, pc2S15, detected StuI and TaqI RFLPs at IDS with heterozygote frequencies of 0.50 and 0.08 respectively. VK16 and VK18 did not detect RFLPs in genomic DNA digested with over 25 different restriction endonucleases.

Multipoint linkage analysis of these polymorphic loci in normal pedigrees indicated that the locus order was F9-(DXS105,DXS98)-(DXS369,VK23)-(VK21,IDS)-DXS304-DXS52. The recombination fractions between adjacent loci were F9-(.058)-DXS105-(.039)-DXS98-(.123)-DXS369-(.00)-VK23-(.057)-VK21-(.00)-IDS-(.012)-DXS304-(.120)-DXS52.

INTRODUCTION

The seminal paper by Botstein et al. (1980) indicated the power of linkage analysis to localize disease genes and polymorphic loci on the human gene map. This approach has been very successful, as indicated by the increasing volume of linkage data presented at the International Human Gene Mapping Workshops (Human Gene Mapping 10, 1989). However this success has not been uniform throughout the human genome.

The development of a precise genetic map around FRAXA has been strikingly slow due to the lack of closely linked polymorphic loci. A panel of cell lines with precisely delineated X chromosome breakpoints was used to physically map a series of probes in relation to DXS369, DXS304, and FRAXA (Chapter 4). The order of probes and loci near FRAXA was:

cen-DXS369,VK24-VK47-VK23-VK16,FRAXA-VK21-VK18-IDS-VK37-DXS304-qter.

Restriction fragment length polymorphisms (RFLPs) have been identified at DXS369 and DXS304 (Oostra et al. 1990; Dahl et al. 1989a). Linkage studies in fragile X pedigrees indicated that these loci lay within 5 centimorgans (cM) of FRAXA (Oostra et al. 1990; Vincent et al. 1989b), but a combined linkage analysis of both these loci has not been presented.

RFLPs at other loci near FRAXA are presented in this Chapter. The probes VK21, VK23, and pc2S15 (IDS) detected RFLPs with a variety of restriction endonucleases. A multipoint linkage analysis of both the established RFLPs and these new polymorphisms in a series of normal CEPH pedigrees is presented.

MATERIAL and METHODS

The DNA methods, method of subcloning, and approach to identifying RFLPs are described in Chapter 2.

Eight DNA probes were used to search for RFLPs near FRAXA:

VK21A was a 1.3 kb HindIII-SalI fragment of single-copy DNA from the VK21. It was subcloned into the HindIII-SalI sites of pBR328. **VK21B** was a 2.2 kb HindIII fragment of single-copy DNA from VK21. **VK21C** was a 2.5 kb HindIII fragment of single-copy DNA from VK21. **VK21D** was a 3.2 kb HindIII fragment of single-copy DNA from VK21. VK21B, VK21C, and VK21D were each subcloned into the HindIII site of pUC18.

VK23B was a 4.4 kb EcoRI fragment of VK23 that was subcloned in pUC19.

VK16B3 was a 0.9 kb HindIII single-copy DNA fragment of VK16 that was subcloned in pSP64 by J Nancarrow (Adelaide, S.A.).

VK18A was a 1.2 kb SalI-HindIII single-copy DNA fragment of VK18 that was subcloned in pBR328 by J Nancarrow.

pc2S15 was a 1.5 kb IDS cDNA clone (Chapter 4).

The candidate performed RFLP searches with the VK21 subclones, VK23B, and pc2S15 using 10-12 restriction endonucleases. Dr S. Yu (Adelaide, S.A.) searched for RFLPs using the VK21 subclones and a further eight restriction endonucleases; she documented the BclI RFLP detected by VK21A. Dr JC Mulley searched for RFLPs using VK23B and a further three restriction endonucleases; he documented the XmnI RFLP detected by VK23B. J Nancarrow performed the RFLP searches using the probes VK16B3 and VK18A.

More than one RFLP was detected by the VK21 subclones, VK23B, and pc2S15. The linkage disequilibrium constants between the RFLPs at a locus were estimated in the manner described by Thompson et al. (1988). The best estimate (D^*) of the linkage disequilibrium constant, D, is given by $D^*=h-pq$ where h is the frequency of the haplotype with both rare alleles, and p and q are the rare allele frequencies. If the two RFLPs are in linkage equilibrium, D equals zero. The following procedure may be used to determine whether D^* is significantly different from zero. The standardized value (D^S) of D^* is given by $D^S=D^*(N^{0.5})/(p(1-p)q(1-q))^{0.5}$ where N is the number of haplotypes examined. The square of D^S is distributed as a X^2 distribution with one degree of freedom.

Forty normal pedigrees from the Centre d'Etude du Polymorphisme Humain (CEPH)(Daussett et al. 1990) were genotyped for the RFLPs listed in Table 5-A. Most of the

Table 5-A
Restriction fragment length polymorphisms used in the linkage study.[a]

Locus	Probe	Enzyme	Alleles	PIC[b]	Location	Reference
F9	pVIII	TaqI	1.8,5/1.3,5 kb	0.33	Xq26.3-q27.1	Kidd et al. 1985
DXS105	cX55.7	TaqI	3.2/4.5 kb	0.11	Xq27.1-q27.2	Kidd et al. 1985
	55E	MspI	16/10 kb	0.48		
DXS98	4D-8	MspI	25/7.8 kb	0.30	Xq27.2	Kidd et al. 1985
DXS369	RN1	XmnI	1.25/1.1 kb	0.48	Xq27.2-q27.3	Oostra et al. 1990
		TaqI	4.5/4.3 kb	0.24		Oberle et al. 1990
DXS297	VK23B	HindIII	10.5/9.5 kb	0.34	Xq27	this chapter
DXS296	VK21A	TaqI	10.9/9.9 kb	0.23	Xq27.3-q28	this chapter
	VK21C	MspI	12.7/9.9 kb	0.31		
IDS	pc2S15	StuI	17.8/15,2.8 kb	0.50	Xq28	this chapter
DXS304	U6.2	TaqI	7.0/3.3 kb	0.36	Xq28	Dahl et al. 1989a
		MspI	4.5/2.2 kb	0.36		
	U6.2-20E	BanI	9.6/5.8 kb	0.49		Rousseau et al. 1990
DXS52	St14-1	TaqI	many	0.80	Xq28	Kidd et al. 1989

Note

a Other RFLPs have been documented at many of the loci. The CEPH pedigrees were not typed for the XmnI RFLP detected by VK23B at DXS297 or for the TaqI RFLP detected by pc2S15 at IDS.

b Polymorphism information content; for X-linked loci this is the frequency of females heterozygous at the locus.

genotype data at F9, DXS105, DXS98, DXS369, DXS304, and DXS52 were provided by Dr I Oberle, (Strasbourg). Linkage studies utilizing some of these data have been published previously (Oberle et al. 1986b, 1987; Arveiler et al. 1988a). Genotype data determined with the 'VK' probes or pc2S15 were obtained by the candidate or by HM Kozman and J McCure under his direction.

Linkage analyses were performed using the LINKAGE programs modified for use with the CEPH three-generation pedigrees (Chapter 2). The A-test (Ott 1985; pp. 105-109; Chapter 2)(as implemented in the computer program HOMOG2) was used to test for homogeneity in the two-point recombination fractions between F9-IDS and IDS-DXS52.

RESULTS

Detection of RFLPs

The four VK21 subclones were used to probe panels of genomic DNA digested with the following restriction endonucleases: BanI, BclI, BglI, BglII, BstNI, BstXI, EcoRI, EcoRV, HincII, HindIII, MspI, PstI, PvuII, RsaI, SacI, TaqI, XbaI, and XmnI.

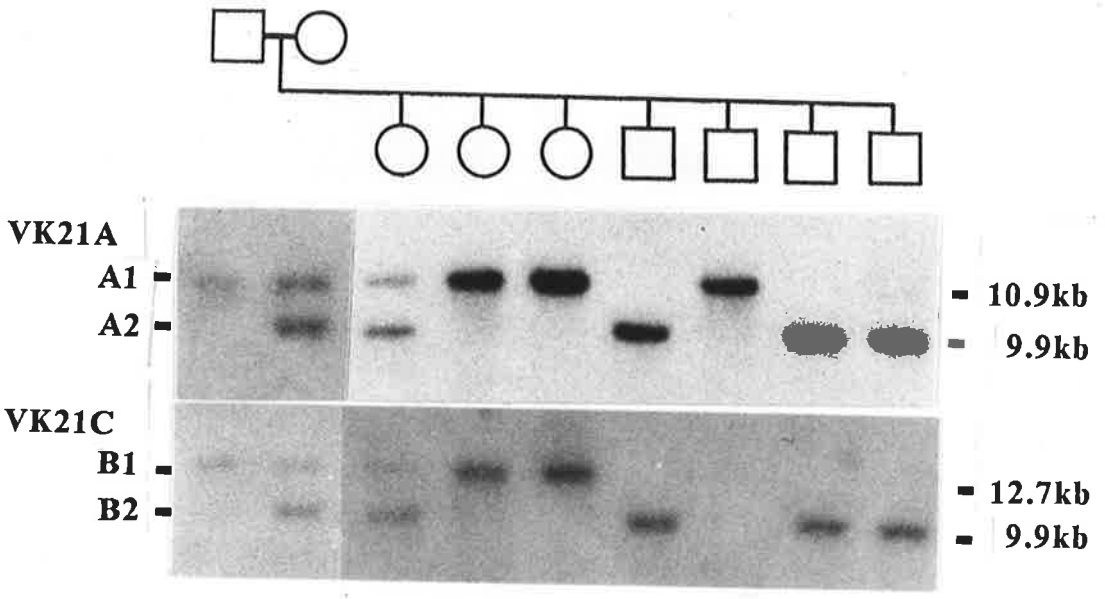
VK21A detected two-allele RFLPs at DXS296 in genomic DNA digested with TaqI (alleles A1,A2)(Fig. 5-1) or BclI (alleles C1,C2)(Yu et al. 1990 [Appendix D]). The heterozygote frequencies for the two RFLPs were 0.23.

VK21C detected a two-allele RFLP at DXS296 in genomic DNA digested with MspI (alleles B1,B2)(Fig. 5-1). The heterozygote frequency was 0.31.

The three RFLPs detected by the VK21 subclones were in strong linkage disequilibrium. Among 45 X chromosomes from 27 unrelated individuals, the haplotypes were either A1B1C1 or A2B2C2. The standardized linkage disequilibrium constants were as follows: $D^S_{AB}=5.36$ ($X^2_1=28.7$; $p<0.001$); $D^S_{AC}=6.71$ ($X^2_1=45.0$; $p<0.001$); and $D^S_{BC}=5.36$ ($X^2_1=28.7$; $p<0.001$).

The probes VK21B and VK21D did not detect an RFLP.

Figure 5-1.
Mendelian inheritance of the TaqI RFLP (A1,A2) detected by VK21A and the MspI RFLP (B1,B2) detected by VK21C at DXS296. For each RFLP the two alleles are indicated on the left of the figure. The sizes of the polymorphic fragments are indicated to the right. Neither probe detected constant bands. Among 111 X chromosomes from unrelated CEPH males and females the A1 allele frequency was 0.87; among 116 X chromosomes from unrelated CEPH males and females the B1 allele frequency was 0.81.



VK23B detected two-allele RFLPs at DXS297 in genomic DNA digested with HindIII (alleles A1,A2) or XmnI (alleles B1,B2) (Fig. 5-2). The heterozygote frequencies for the two RFLPs were 0.34 and 0.49 respectively. The two RFLPs were in linkage disequilibrium. Among 19 DNA samples from unrelated Caucasian males the haplotype frequencies were A1B1 (0.42), A1B2 (0.21), A2B1(0.37), and A2B2 (0.0). The standardized linkage disequilibrium constant (D^S) was -2.01 (X^2_1=4.04; $p<0.05$). VK23B did not detect RFLPs in genomic DNA digested with any of the following enzymes: AvaI, AvaII, BamHI, BclI, BglII, EcoRI, HincII, MspI, PstI, PvuII, Sau3A, StuI, and TaqI.

VK16B3 did not detect an RFLP at DXS293 in genomic DNA digested with the following enzymes: AluI, AvaI, AvaII, BamHI, BanI, BclI, BglI, BglII, BstNI, BstXI, DdeI, DraI, EcoRI, EcoRV, HaeIII, HincII, HindIII, HinfI, MboII, MspI, NsiI, PstI, PvuII, RsaI, SacI, Sau3A, StuI, TaqI, XbaI, and XmnI.

VK18A did not detect an RFLP at DXS295 in genomic DNA digested with the following enzymes: AvaI, AvaII, BamHI, BanI, BanII, BclI, BglI, BglII, BstNI, BstXI, DraI, Eco0109, EcoRI, EcoRV, HaeIII, HincII, HindIII, HinfI, MboII, PstI, PvuII, RsaI, SacI, StuI, TaqI, XbaI, and XmnI.

Figure 5-2 (a).
Mendelian inheritance of the HindIII RFLP detected by VK23B at DXS297. This is a composite figure made from two autoradiographs. DNA size markers indicate the DNA fragment sizes in the two parental lanes on the left of the figure; the two alleles are indicated in the lanes of the offspring by arrows to the right of the figure. The polymorphic fragments were 10.5 kb (A1 allele) and 9.5 kb (A2 allele) long. A faint constant 6.7 kb fragment was also detected. Some lane background was noted when using VK23B, but this did not obscure the polymorphic bands. Among 117 X chromosomes from unrelated CEPH males and females the A1 allele frequency was 0.78.

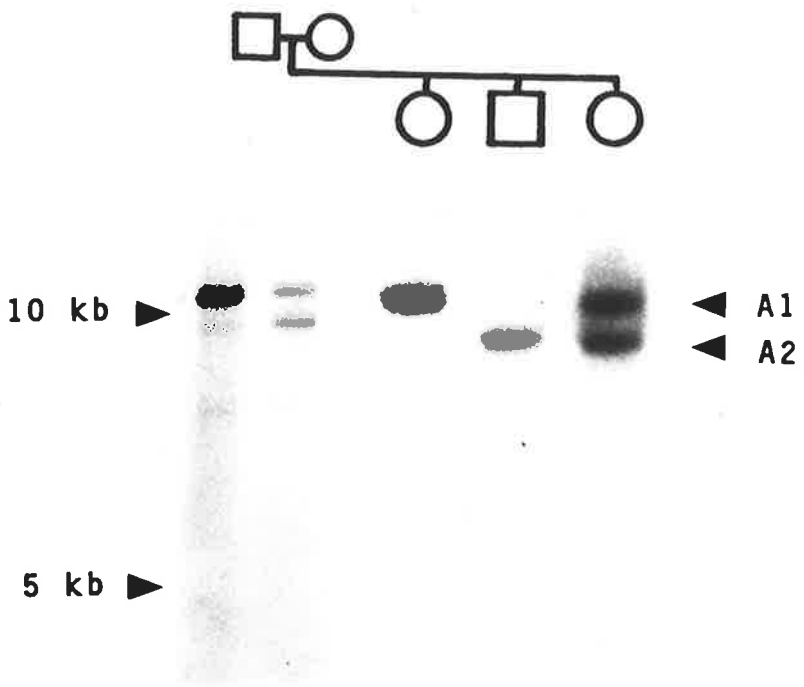
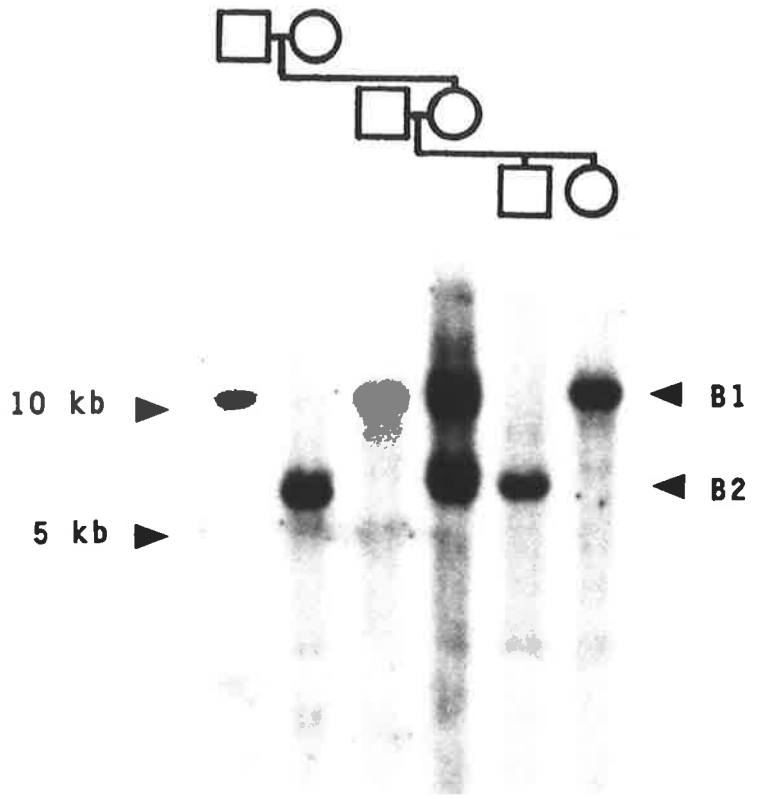


Figure 5-2 (b).
Mendelian inheritance of the XmnI RFLP detected by VK23B at DXS297. DNA size markers are indicated on the left of the figure; the alleles are indicated on the right. The polymorphic fragments were 10.3 kb (B1 allele) and 6.6 kb (B2 allele) long. No constant bands were detected. Among 42 X chromosomes from unrelated Caucasian males and females the B1 allele frequency was 0.57.



The probe pc2S15 detected two-allele RFLPs at IDS in DNA digested with StuI (alleles A1,A2) or TaqI (alleles B1,B2) (Fig. 5-3). The heterozygote frequencies for the two RFLPs were 0.50 and 0.08 respectively. There was no evidence that the two RFLPs were in linkage disequilibrium. Among 27 X chromosomes from unrelated CEPH males and females the haplotype frequencies were A1B1 (0.60), A1B2 (0.04), A2B1 (0.36), and A2B2 (0.0). The standardized linkage disequilibrium constant (D^S) was -0.959 (X^2_1=0.92; p>0.1). RFLPs were not detected with the following enzymes: BamHI, BclI, BglII, EcoRI, HincII, HindIII, MspI, PstI, PvuII, and Sau3A.

Linkage Analysis

The CEPH pedigrees were genotyped for RFLPs listed in Table 5-A. Where the CEPH pedigrees were genotyped for more than one RFLP at a locus, the genotypes were reduced to a single pair of alleles by hand (Chapter 2). The data were then checked with program XPHASE (Chapter 2; Appendix A) and by hand. No double recombinants in the interval F9-DXS52 were observed. All these genotype data have been communicated to the CEPH database and will be included in the CEPH consortium linkage map of the X chromosome. The pedigrees and the genotypes are listed in Appendix B.

Two-point LOD scores and recombination fractions for all pair-wise combinations of loci are summarized in Table 5-B.

Figure 5-3 (a).
Mendelian inheritance of the StuI RFLP detected by pc2S15 at IDS. This is a composite figure made from two autoradiographs. The DNA size markers indicate the DNA fragment sizes in the two parental lanes on the left of the figure; the two alleles are indicated in the lanes of the offspring by arrows to the right of the figure. The polymorphic fragments were 17.8 kb (A1 allele), and 15.0 and 2.8 kb (A2 allele) long. Constant fragments of the following sizes were also detected: 2.6, 3.2, 3.4, 3.6, and 4.7 kb. Among 104 X chromosomes from unrelated CEPH males and females the A1 allele frequency was 0.55.

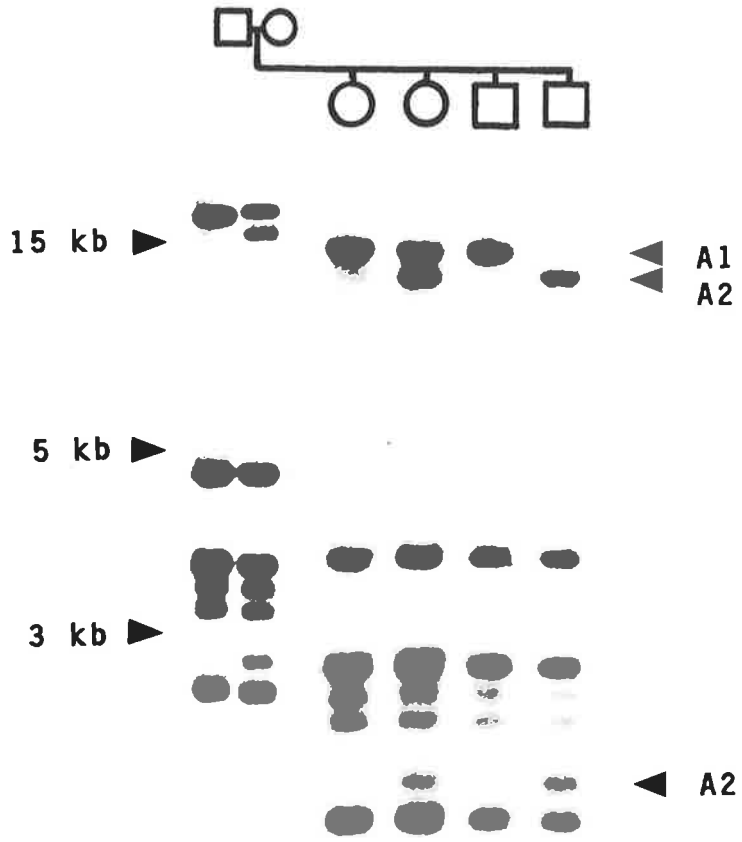
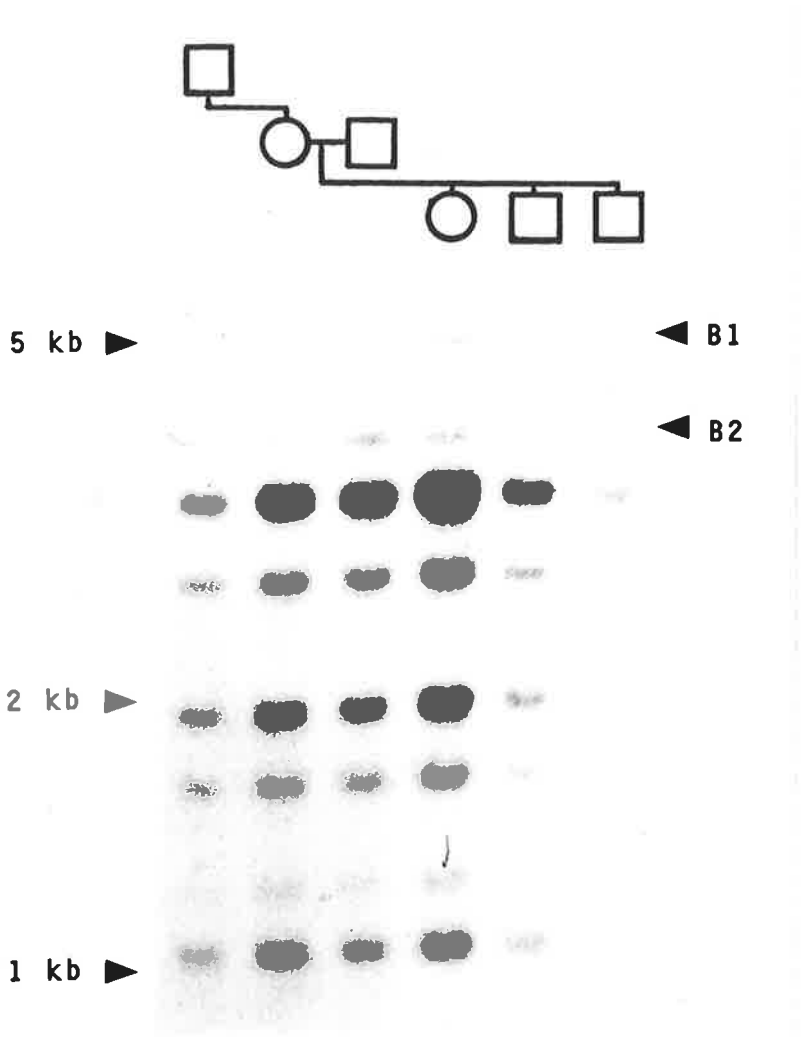


Figure 5-3 (b).
Mendelian inheritance of the TaqI RFLP detected by pc2S15 at IDS. DNA size markers are indicated on the left of the figure; the alleles are indicated on the right. The polymorphic fragments were 5.1 kb (B1 allele) and 3.8 kb (B2 allele) long. Constant fragments of the following sizes were also detected: 1.1, 1.3, 1.7, 2.0, 2.8, and 3.3 kb. The polymorphic bands were much fainter than the constant bands. Among 27 X chromosomes from unrelated CEPH males and females the B1 allele frequency was 0.96.



Note to Table 5-B

Two-point LOD scores were calculated for each pair of loci. The figures above the diagonal are the best estimate of the recombination fraction (upper figure) and the peak LOD score (lower figure). The figures below the diagonal are the standard error of the recombination fraction (upper figure) and the number of pedigrees informative for each pair of loci (lower figure). The standard error of the recombination fraction is the square root of the variance estimated by the LINKAGE programs. (NC not calculated).

Table 5-B

Summary of two-point linkage analysis in the CEPH pedigrees.

	F9	DXS105	DXS98	DXS369	VK23	VK21	IDS	DXS304	DXS52
F9		.043 14.34	.078 6.38	.221 2.54	.251 1.36	.190 1.28	.193 4.88	.241 2.62	.327 1.46
DXS105	.026 9		0 6.62	.170 6.79	.179 1.38	.142 2.28	.209 4.77	.170 5.50	.279 3.55
DXS98	.044 6	NC 3		0 2.71	.072 2.65	.152 1.19	.125 4.59	.069 5.27	.237 1.81
DXS369	.060 6	.044 9	NC 1		0 4.52	.063 6.08	.054 15.19	.062 12.74	.124 10.83
VK23	.089 3	.094 2	.072 2	NC 2		.155 1.19	.077 6.85	.042 5.12	.175 2.36
VK21	.087 4	.078 3	.103 2	.045 5	.105 2		0 15.95	0 7.53	.154 5.06
IDS	.049 10	.049 10	.054 6	.028 10	.045 5	NC 8		.014 18.20	.136 13.93
DXS304	.062 7	.050 8	.049 4	.031 8	.043 3	NC 4	.017 9		.127 11.25
DXS52	.056 12	.048 13	.076 6	.036 11	.074 5	.051 9	.032 17	.036 12	

The peak LOD scores for the recombination fractions between loci that were adjacent on the basis of physical mapping ranged from 1.19 (VK23-VK21) to 18.20 (IDS-DXS304). No recombinants were observed between DXS105-DXS98, DXS98-DXS369, DXS369-VK23, and VK21-IDS.

For multipoint linkage analysis, the order of loci determined by physical mapping (Chapter 4) was assumed, and the relative likelihood of adjacent pairs being inverted was calculated (Fig. 5-4). In general it was not possible to order adjacent pairs of loci where there was no recombination on two-point linkage analysis. The one exception was DXS98-DXS369 where DXS98 was placed proximal to DXS369. Although no recombination was observed between these two loci, DXS369 demonstrated much closer linkage to distal markers than did DXS98 (Table 5-B). The odds in favor of placing DXS304 distal to IDS were only 22:1. There was only one recombination event between IDS and DXS304, and this was in a pedigree uninformative both for VK21 and at DXS52.

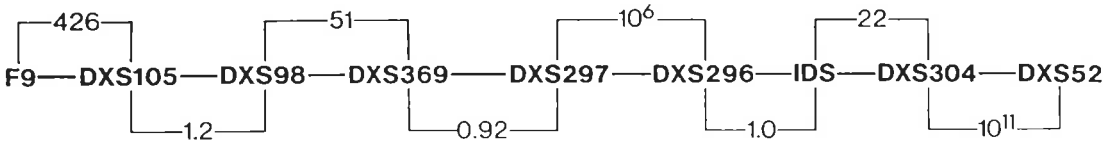
The order of loci suggested by this analysis was F9-(DXS105,DXS98)-(DXS369,VK23)-(VK21,IDS)-DXS304-DXS52. As an aid for ordering other loci in this region, individuals who were recombinant between loci in the interval DXS369-DXS304 are listed in Table 5-C. Assuming that DXS98 was distal to DXS105 (Keats et al. 1989; Chapter 4) the recombination

Figure 5-4.

Multipoint linkage analysis of nine loci at Xq27-q28. The probes which detect RFLPs at the loci shown are listed at the bottom of the figure.

(Upper) The order of loci was derived from physical mapping studies (Chapter 4). The odds against inverting adjacent loci were calculated using genotypes of the CEPH pedigrees and the program ILINK.

(Lower) Recombination fractions between adjacent loci were estimated in a multipoint analysis of the most likely order.



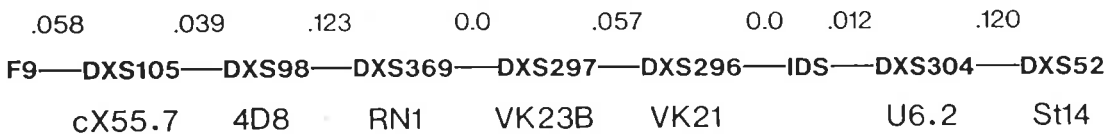


Table 5-C

Offspring in the CEPH pedigrees with phase-known recombination between loci in the interval DXS369-DXS304.

Closest informative loci	Pedigree number (individual)
DXS369:DXS297	none
:DXS296	13292 (8); 1362 (11).
:IDS	1416 (8); 1421 (12).
:DXS304	884 (14).
DXS297:DXS296	1344 (5, 11)[a].
:IDS	1408 (7).
:DXS304	none
DXS296:IDS	none
:DXS304	none
IDS :DXS304	1416 (6).

Note

a phase inferred.

fractions between adjacent loci were derived by multipoint linkage analysis and are shown in Fig. 5-4.

Homogeneity testing

The A-test was used to determine whether the pair-wise LOD scores for the recombination fractions between F9-IDS and IDS-DXS52 were heterogeneous. The locus IDS was chosen because more CEPH pedigrees were informative at this locus than at other nearby loci. For F9-IDS there were 10 informative pedigrees with 45 phase-known and 27 phase-unknown meioses. There was insufficient evidence to indicate linkage heterogeneity (X^2_2=0.078; p>0.2). For IDS-DXS52 there were 17 informative pedigrees with 82 phase-known and 39 phase-unknown meioses. Again there was insufficient evidence to indicate linkage heterogeneity (X^2_2=2.840; p>0.1).

DISCUSSION

Linkage analysis of RFLPs at Xq27-q28

RFLPs at three loci near FRAXA are described in this Chapter. These RFLPs will be of great value in genetic counseling of families with the fragile X syndrome or Hunter syndrome (IDS deficiency). The RFLPs detected by the VK21 subclones at DXS296 are in linkage disequilibrium, as are the RFLPs detected by VK23B at DXS297. Although there was no evidence of linkage disequilibrium between the RFLPs at IDS, the power of this assessment to exclude linkage disequilibrium was low (Thompson et al. 1988).

The RFLPs defined by VK23 and VK21 flank FRAXA, and are closer to FRAXA than DXS369 or DXS304 (Fig. 5-4; Chapter 4). A linkage analysis of the RFLPs detected by these probes in fragile X families is presented in the next chapter.

With regard to linkage studies of Hunter syndrome, RFLPs have now been defined at the gene itself and at the flanking loci, DXS296 (VK21) and DXS304. The genetic distances between IDS and these loci are small (Figure 5-4) and have narrow confidence intervals. A proportion of affected males have deletions at IDS which are detected by pc2S15 (Wilson et al. 1990 [Appendix D]), and the identification of carriers within a family may be relatively simple. In families with an established mutation where a deletion is not evident, it is

now possible to provide carrier risk estimates utilizing RFLPs at IDS and at the closely linked flanking loci.

The linkage relationships of nine loci at Xq27-q28 were also defined in this study. The order of loci determined by linkage analysis was consistent with that determined by physical mapping studies (Chapter 4). The physical mapping studies had been based on the assumption that the cell lines used did not have complex rearrangements at Xq27-q28. The observations that CY34A had an interstitial deletion at Xq26 and that the human chromosomal content of APC-5 was fragmented (Chapter 4) suggests that such rearrangements could occur. Therefore it is reassuring that the results of this linkage study were compatible with the order based on physical mapping studies.

However the limitations of linkage analysis for ordering loci were evident. Despite genotyping 40 three-generation pedigrees at nine loci there were three pairs of loci that could not be ordered on the basis of linkage analysis, DXS105-DXS98, DXS369-VK23, and VK21-IDS. There are three reasons for this. First, the RFLPs tested at DXS98 and DXS297 (VK23) were informative in only a few pedigrees. Second, recombination between X-linked loci occurs in only one parent, thus reducing the information that can be derived from a study of X-linked rather than autosomal loci. Third, as more loci are identified in a region it becomes

increasingly less likely that recombination will be observed between an adjacent pair of loci. The order of loci shown in Fig. 5-4 is based on both physical mapping studies (Chapter 4) and the linkage analysis presented in this chapter. There is some independent evidence to support this order. Collated linkage data has placed DXS98 distal to DXS105 (Keats et al. 1989), and IDS has been placed distal to VK21 in pulsed field gel electrophoresis studies (Hyland et al., manuscript submitted). The location of DXS297 distal to VK23 (Chapter 4) has not been independently confirmed.

With the development of a large scale restriction map of the region close to FRAXA (Hyland et al., manuscript submitted) it is possible to correlate physical and genetic distances. IDS and DXS304 are within 900 kb of each other, and the recombination fraction between them was 0.012. VK21 lies approximately 800 kb proximal to IDS (Yu S., personal communication) and no recombinants were observed between these loci. These figures are consistent with the genome average of 1 cM per 1000 kb of DNA. There is no evidence of a recombination 'hot-spot' at distal Xq27[1].

---

1 Xq27 has been proposed as a region of preferential recombination in normal and fragile X families. This proposal is analysed and discussed in Chapter 7.

Linkage homogeneity in CEPH pedigrees

These linkage studies were performed under the assumption that there was no heterogeneity in the linkage relationships of co-dominant loci. This assumption need not be true. Some evidence has been presented that there is linkage heterogeneity between F9 and DXS52 in fragile X pedigrees (Brown et al. 1985, 1986, 1987b, 1988; Risch 1988). Brown et al. (1987b) also reported definite evidence of linkage heterogeneity between F9 and FRAXA, but this has been disputed (Clayton et al. 1988). There is no evidence of linkage heterogeneity between FRAXA and DXS52 in fragile X pedigrees (Brown et al. 1987b; Risch 1988). Linkage heterogeneity could reflect a sampling fluctuation, or be specific for the fragile X mutation, or also be a feature of normal pedigrees. If linkage heterogeneity is a feature of normal pedigrees, it may be necessary to utilize more than 40 CEPH pedigrees to develop a representative linkage map of the human genome.

It is impossible to test for linkage heterogeneity in the interval F9-FRAXA in normal pedigrees. IDS was chosen as an appropriate locus for this test for two reasons. First, a large number of the CEPH pedigrees were informative at IDS. Second, IDS must be close to FRAXA; IDS lies between FRAXA and DXS304 (Chapter 4) which are separated by a genetic distance of approximately 3 cM (Vincent et al. 1989b). No

evidence of linkage heterogeneity was found between F9-IDS and IDS-DXS52 in this study of normal pedigrees.

It is difficult to estimate the power of this study to exclude linkage heterogeneity. In the study of Brown et al. (1987b) 24% of fragile X pedigrees showed no recombination between F9-FRAXA while the remainder demonstrated a recombination fraction of 0.37. Ott (1986) and Risch (1988) have tabulated the power of heterogeneity tests for a variety of alternative hypotheses, but they do not consider the combination of values noted by Brown et al. (1987b). Cavalli-Sforza and King (1986) suggested that 24 phase-known pedigrees each with four children were sufficient to detect linkage heterogeneity at odds of 10:1 when 20% of the pedigrees were tightly linked and the remainder were unlinked. In this study the pedigrees all had more than four children, thus increasing the power of the study (Ott, 1985; p. 54). On the other hand, under the alternative hypothesis that 80% of the pedigrees were loosely linked rather than unlinked, the power of the heterogeneity test would be reduced (Risch 1988).

Conclusion

The development of this genetic map at Xq27-q28 will be crucial for the precise localization of disease loci such as FRAXA in this region of the X chromosome. Such a map also may be the basis for estimating approximate confidence intervals

for gene location, and for providing reliable carrier risk estimates to consultands as discussed in Chapter 3.

Chapter 6.

GENETIC MAPPING OF NEW RFLPS IN FRAGILE X FAMILIES DEFINES A STRATEGY FOR DNA STUDIES IN THE FRAGILE X SYNDROME.

SUMMARY

The unusual genetics and cytogenetics of the fragile X syndrome make genetic counselling difficult. DNA studies have been of limited value in genetic counselling because the nearest polymorphic DNA loci had recombination fractions of 12% or more with FRAXA.

Five polymorphic loci have recently been described in this region of the X chromosome, DXS369, VK23, VK21, IDS, and DXS304. The positions of these loci in relation to FRAXA were defined in a collaborative linkage study of 112 affected families. On multipoint linkage analysis the five loci had recombination fractions of 4% or less with FRAXA. The closest locus, VK21 was 2.2 cM distal to FRAXA.

The polymorphisms at these loci can be detected in DNA digested with a limited number of restriction endonucleases. A strategy for DNA studies which is based on three restriction endonucleases and five probes will detect one or more of these polymorphisms in 94% of women. This strategy greatly increases the utility of DNA studies in providing genetic advice to families with the fragile X syndrome.

INTRODUCTION

Once a child has been diagnosed as having the fragile X syndrome, other members of the family frequently seek genetic counselling, either to determine their risk of being a carrier or to request prenatal diagnosis (Turner et al. 1986). The unusual pattern of inheritance of the syndrome and the lack of an unequivocal marker for those carrying the mutation have made it difficult to provide reliable risk estimates.

It is possible to determine an individual's risk of being a carrier by observing the inheritance of DNA polymorphisms located near FRAXA (Sutherland & Mulley 1990). In general, the value of a DNA polymorphism for providing genetic counseling in a disorder is determined by two factors (Gusella 1986). The first is the genetic distance between the polymorphism and the disease gene.

The polymorphic loci generally used in studies of fragile X families have been F9, DXS105, DXS98, and DXS52 (Sutherland & Mulley 1990). Each of these loci has a recombination fraction of more than 0.12 with FRAXA (Mandel et al. 1989). Consequently, identification of a fragile X carrier on the basis of the inheritance of one of these polymorphisms would be incorrect in over 12% of cases.

In view of the large recombination fractions between the nearest polymorphic loci and FRAXA, an estimate of carrier risk was ideally based on the inheritance of any two polymorphisms which flank FRAXA. In this situation the risk of incorrectly identifying a carrier may be as low the product of the two recombination fractions (i.e. 0.014). However, the advantage of using flanking polymorphisms is lost if recombination has occurred in the interval between the two polymorphic loci, making it impossible to determine which of the two polymorphisms was inherited with FRAXA. The probability that such a recombination will occur between two loci is approximately equal to the sum of the recombination fractions (i.e. at least 0.24).

The second consideration in choosing a DNA polymorphism is whether the polymorphism is informative in the family being investigated. The ideal polymorphic locus for investigating families with the fragile X syndrome would be very close to FRAXA and have a high probability of being informative in the women of the family. In practice it is usually necessary to examine the family at a number of polymorphic loci, and to seek some compromise between testing highly polymorphic loci far from FRAXA and testing closer but relatively uninformative loci. Strategies for DNA studies in fragile X families have been proposed (Heilig et al. 1988; Sutherland & Mulley 1990) which indicate the order in which the

polymorphisms should be evaluated to provide the most efficient diagnostic service for the genetic counselor.

Two polymorphic loci, DXS369 (Oostra et al. 1990), and DXS304 (Vincent et al. 1989b), have been reported to have recombination fractions of less than 0.05 with FRAXA. A combined linkage study of these two loci in fragile X families has not been reported.

A further three polymorphic loci, DXS296 (probes VK21A and VK21C), DXS297 (VK23B), and IDS (pc2S15), have been localized to this region on the basis of physical mapping (Chapter 4) and genetic mapping in normal pedigrees (Chapter 5). The order of loci is cen-DXS369-VK23-FRAXA-VK21-IDS-DXS304-qter.

A pilot linkage study of the location of VK21 in seven fragile X families (Suthers et al. 1989a [Appendix D]) had suggested that VK21 was the closest probe to FRAXA. The VK21 subclones were distributed to diagnostic laboratories around the world, and a collaborative study of the linkage relationships of these probes and FRAXA was initiated.

In a multipoint linkage study of 112 fragile X families, the five loci all had recombination fractions of 0.04 or less with FRAXA. The closest probe was VK21 which had a recombination fraction of 0.02 with FRAXA. Over 98% of women were heterozygous at one or more of these loci. On the basis

of these results, an efficient strategy for DNA studies in families with the fragile X syndrome is presented.

METHODS

Selection of pedigrees

The collaboration was initiated and managed by the candidate. Data from a total of 153 families were obtained from 13 centers around the world. As affected individuals with no affected relatives could represent new mutations (Sherman et al. 1988b), families were included in the analysis only if at least one family member was mentally retarded and expressed the fragile X, and provided at least one further family member expressed the fragile X. Expression of the fragile X was assessed by culturing peripheral lymphocytes under specific conditions (Sutherland 1990). Expression of the fragile X in 1% or more of lymphocytes was regarded as positive.

Among the families contributed from overseas laboratories, there was a general selection bias in favor of families having women heterozygous at loci near FRAXA. There was also a specific bias in favor of families with women heterozygous at DXS296 (VK21). Details regarding the pedigrees are listed in Appendix C.

Linkage analysis

Various DNA probes were used to identify restriction fragment length polymorphisms (RFLPs) at nine loci near FRAXA. The approximate positions of the loci on the X chromosome are

shown in Figure 6-1. Details of the RFLPs are summarized in Table 6-A.

All the genotype and pedigree data were checked by hand. Pedigrees having a single affected individual or with apparent non-Mendelian inheritance of a polymorphism were excluded. A total of 1368 individuals from 112 pedigrees were included in the analysis. The pedigree and genotype data are listed in Appendix C. The number of families informative at each locus is shown in Table 6-B.

Details of the LINKAGE programs and the choice of parameters for the analysis are given in Chapter 2. The program LINKMAP was used for the multipoint linkage analysis, and FRAXA was localized in relation to the genetic map defined in normal pedigrees (Chapter 5). As detailed in Chapter 2, the background genetic map consisted of the positions of DXS98, DXS369, VK23, VK21, IDS, and DXS304. Genotype data at one or more of these loci were available from 101 of the fragile X pedigrees.

An approximate 95% confidence interval for FRAXA location was determined using the protocol outlined in Chapter 2 and the program BOOTMAP [Appendix A].

Figure 6-1.
The order of polymorphic loci down the X chromosome is indicated. VK23 detects the locus DXS297, and VK21 detects DXS296. The dark band on the ideogram corresponds to the band at Xq27 noted after G-banding of the chromosome. The positions of the loci and of the fragile site in relation to the chromosome bands are approximate.

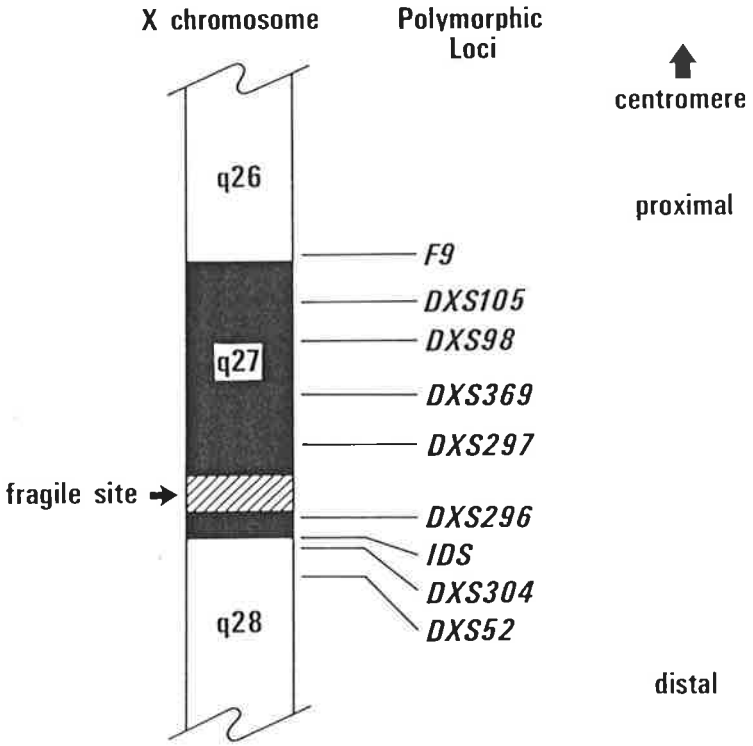


Table 6-A
Restriction fragment length polymorphisms used in the linkage study.

Locus	Probe	Enzyme[a]	Heterozygosity[b]	Location	Reference
F9	pVIII	TaqI	0.38	Xq26.3-q27.1	Kidd et al. 1989
DXS105[c]	cX55.7	TaqI	0.11	Xq27.1-q27.2	Kidd et al. 1989
		HindIII	0.48		
		EcoRI	0.40		
		StuI	0.41		
	55E	PstI	0.48		
DXS98[c]	4D-8	MspI	0.30	Xq27.2	Kidd et al. 1989
		XmnI	0.08		
	4D8-B	BglII	0.15		Schnur et al. 1989
	4D-8IV	XmnI	0.08		
DXS369[c]	RN1	XmnI	0.48	Xq27.2-q27.3	Oostra et al. 1990
		TaqI	0.24		Oberle et al. 1990
DXS297[c]	VK23B	HindIII	0.34	Xq27	Chapter 5
		XmnI	0.49		
DXS296[c]	VK21A	TaqI	0.23	Xq27.3-q28	Chapter 5
		BclI	0.23		
	VK21C	MspI	0.31		
IDS	pc2S15	StuI	0.50	Xq28	Chapter 5
		TaqI	0.08		
DXS304[c]	U6.2	many	0.36	Xq28	Dahl et al. 1989a
	U6.2-20E	BanI	0.49		Rousseau et al. 1990
DXS52	St14-1	TaqI	0.80	Xq28	Kidd et al. 1989
		BclI	0.32		

Note

a restriction endonuclease used to digest DNA samples from family members.

b frequency of women heterozygous at the locus.

c the RFLPs at this locus are in linkage disequilibrium, i.e. women who are not informative for one RFLP have a reduced probability of being informative for other RFLPs at the locus.

RESULTS

Two-point linkage analysis

The results of two-point linkage analysis of FRAXA and each of the nine polymorphic loci are summarized in Table 6-B. Recombination was observed between FRAXA and each of the loci. The pedigrees demonstrating recombination are identified in Appendix C.

The probe closest to FRAXA was VK21 which had a peak LOD score of 33.45 at a recombination fraction of 0.015. This analysis incorporated the data from the pilot study of seven families (Suthers et al. 1989a [Appendix D]). Recombination between VK21 and FRAXA was documented in three affected males from three different families. The adjacent locus IDS (pc2S15) had a recombination fraction of 0.089 with FRAXA. The other locus distal to FRAXA, DXS304, had a recombination fraction of 0.031.

The proximal loci DXS369 and VK23 had recombination fractions with FRAXA of 0.066 and 0.042 respectively.

Multipoint linkage analysis

Multipoint LOD scores were calculated for various positions of FRAXA along the genetic map (Fig. 6-2). The peak multipoint LOD score was 48.49. The corresponding location of FRAXA was 2.2 cM proximal to VK21.

Table 6-B

Summary of two-point linkage analysis of the Fragile X locus, FRAXA, and nearby loci.

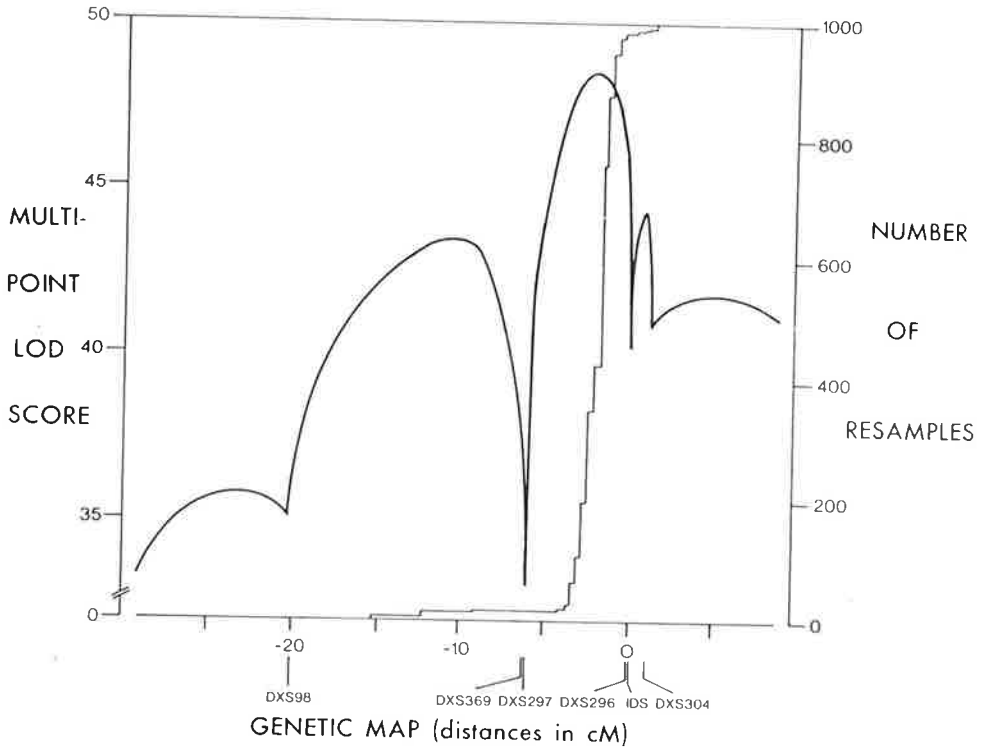
	Recombination fractions										
	0.00	0.01	0.05	0.10	0.20	0.30	0.40	LOD_{max}	R.F.[a]	C.I.[b]	n[c]
FRAXA vs.											
F9	-26.29	-8.85	1.02	4.88	6.49	5.18	2.66	**6.49**	.190	(.12-.28)	44
DXS105	-19.68	-5.11	2.55	4.88	5.11	3.38	1.22	**5.40**	.152	(.08-.24)	37
DXS98	3.65	5.88	6.69	6.47	5.10	3.25	1.36	**6.70**	.058	(.01-.15)	25
DXS369	5.04	11.31	14.50	14.24	11.06	6.74	2.56	**14.62**	.066	(.04-.12)	45
VK23	3.13	4.17	4.46	4.23	3.28	1.99	0.68	**4.46**	.042	(.005-.18)	12[d]
VK21	30.75	33.36	32.05	28.58	20.29	11.56	4.08	**33.45**	.015	(.005-.04)	67
IDS	0.34	2.69	4.01	4.18	3.50	2.32	1.02	**4.19**	.089	(.02-.23)	16[d]
DXS304	7.57	9.30	9.56	8.74	6.27	3.49	1.09	**9.67**	.031	(.005-.10)	27
DXS52	-21.42	2.31	16.84	21.07	19.43	12.98	5.43	**21.45**	.126	(.09-.17)	89

Note
a recombination fraction corresponding to the maximum LOD score.
b approximate 90% confidence interval for the recombination fraction (Conneally et al. 1985).
c number of families polymorphic at the locus.
d this locus was analysed in the Adelaide families only.

Figure 6-2.

Multipoint LOD scores for the location of FRAXA are plotted against genetic location along the X chromosome. The origin of the map was arbitrarily placed at VK21 (DXS296). Distances along the map were derived from recombination fractions (Chapter 5) using Haldane's formula. DXS98 lay 20.2 cM proximal to VK21, DXS369 and VK23 (DXS297) lay 6.1 cM proximal to VK21, IDS was placed coincident with VK21, and DXS304 lay 1.2 cM distal to VK21. Multipoint LOD scores for FRAXA location were calculated at 20 points in the interval between VK23 and VK21; calculations were performed at 5 points in each of the other intervals. The peak of the multipoint LOD score curve (heavy line) occurred 2.2 cM proximal to VK21, indicating that this was the most likely location of FRAXA.

A cumulative histogram of the distribution of FRAXA locations obtained from 1000 resamplings is shown as a fine line. The median location was 2.2 cM proximal to VK21. 95% of the locations (centered at the median) were 0.7 to 3.7 cM proximal to VK21.



In using the program BOOTMAP to determine an approximate confidence interval for FRAXA location, the sets of multipoint LOD scores were resampled in two ways. The first approach was conservative, and the 101 sets of LOD scores were randomly resampled (with replacement). The distribution of the resampled locations of FRAXA under this conservative resampling scheme is shown in Fig. 6-2. The approximate 95% confidence interval for FRAXA location was 0.7 to 3.7 cM proximal to VK21.

The second approach was non-conservative, and was similar to that used in Chapter 3. The resampling was random with the exception that each resample had to include data from at least one family having recombination between FRAXA and each of the loci on the background map. Under this condition, the approximate 95% confidence interval for FRAXA location was 1.3 to 3.7 cM proximal to VK21.

The 'one-LOD-unit-down' confidence interval (Keats et al. 1989) for FRAXA location was 1.0 to 3.7 cM proximal to VK21.

DISCUSSION

Two-point linkage analysis of *FRAXA*.

This collaborative linkage study documented the genetic locations of nine polymorphic loci in relation to FRAXA. The relatively large recombination fractions noted between FRAXA and F9, DXS105, and DXS52 (Table 6-B) are similar to published values (Keats et al. 1989). The recombination fraction of 0.058 between FRAXA and DXS98 is similar to the value initially reported (Brown et al. 1987a), but subsequent pooled studies have indicated that the recombination fraction is more likely to be 0.15 (Mandel et al. 1989). The larger value is also more consistent with the relative positions of polymorphic loci in normal pedigrees (Chapter 5).

Two-point linkage analysis of FRAXA with DXS369, VK23, VK21, IDS, and DXS304 indicated that the best estimates of the recombination fractions were all less than 10%. The recombination fractions between FRAXA and DXS369 and DXS304 were consistent with published values (Oostra et al. 1990; Vincent et al. 1989b).

The recombination fraction between FRAXA and IDS was estimated to be 0.089. This value seems inconsistent with other data. Physical mapping studies have indicated that IDS lies between VK21 and DXS304 (Chapter 4). However, these two loci had recombination fractions with FRAXA of only 0.015 and

0.031 respectively (Table 6-B). This apparent discrepancy is probably not of biological significance, but simply reflects a sampling fluctuation. IDS was studied in a small number of families and the confidence interval for the recombination fraction was wide. Estimates of the recombination fractions between DXS98, DXS369, VK23, VK21, IDS, and DXS304 determined by multipoint linkage analysis do not differ in normal versus fragile X families (Chapter 7). In normal families there was no recombination between IDS and VK21 (Chapter 5), and the true recombination fraction between FRAXA and IDS is likely to be less than 8.9%.

Multipoint linkage analysis of *FRAXA*.

Multipoint linkage analysis is statistically more efficient than two-point linkage analyses, and provides a more accurate and precise genetic map (Lathrop et al. 1985). Estimates of the recombination fractions between the various loci and FRAXA were derived from Fig. 6-B and are summarized in Table 6-C. Multipoint linkage analysis indicated that DXS369, VK23, VK21, IDS, and DXS304 all had recombination fractions of 0.04 or less with FRAXA.

The approximate 95% confidence interval for FRAXA location was very narrow, and reflected the number of families included in the analysis. In applying this method to linkage data from a single pedigree (Chapter 3), it was necessary to use a non-conservative resampling strategy to obtain

appropriate estimates of gene location. The resampling process was also tedious, and it was necessary to perform repeated LINKMAP analyses. These constraints did not apply when estimating a confidence interval for FRAXA location with linkage data from many pedigrees. The difference in confidence intervals derived by conservative versus non-conservative approaches was minimal, and the resampling was performed 1000 times in a matter of minutes.

Defining a recombination fraction of 0.02 between a locus and FRAXA could only be achieved with a collaborative study. Ideally a recombination fraction of this magnitude could be documented by analyzing the DNA of approximately 50 offspring in fragile X pedigrees. However, the presence of unaffected carriers for the fragile X syndrome and the irregular size and structure of human families markedly reduces the amount of information that can be obtained from linkage studies (Ott 1985, p.54, p.132). There has only been one other multipoint linkage study of the fragile X syndrome of this magnitude. Brown et al. (1988) described a linkage study of 147 families. The closest polymorphic loci that were localized in that study were F9 and DXS52, each of which have recombination fractions of over 0.12 with FRAXA.

Strategy for DNA studies in fragile X families.

The definition of five polymorphic loci with recombination fractions of 0.04 or less with FRAXA represents a major

advance in the development of the genetic map in this region of the X chromosome. These polymorphisms have immediate application in genetic counselling. An estimate of genetic risk based on the inheritance of any one of these polymorphic loci would be correct in at least 96% of cases. The inclusion of other pedigree or cytogenetic data in the analysis may reduce the risk even further (Mulley et al. 1987; Sutherland & Mulley 1990).

Less than 50% of women are heterozygous at each of the five polymorphic loci close to FRAXA (Table 6-A), and at first glance these RFLPs might appear to be of little added value in studies of fragile X families. However, two factors argue against such a pessimistic conclusion. First, all five loci are close to FRAXA, and an accurate estimate of carrier risk can be made on the basis of the inheritance of just one polymorphism. The probability that a woman would be heterozygous for at least one of the loci is high. Second, a number of the RFLPs can be detected using the same restriction endonuclease to digest the DNA of family members. The nylon membrane to which the digested DNA is transferred can be re-probed for a number of different RFLPs, and it is possible to rapidly screen the RFLPs that are close to FRAXA.

An efficient strategy for DNA studies in families with the fragile X syndrome is presented in Table 6-C. Step 1 involves digesting the DNA samples of family members with three

Table 6-C.

Strategy for DNA studies of fragile X families

	Digest DNA with	Probe DNA with		Recombination fraction with FRAXA
STEP 1				
	TaqI[a]	VK21A	(DXS296)	0.02 distal[b]
	StuI	pc2S15	(IDS)	0.02 distal
	XmnI	VK23B	(DXS297)	0.04 proximal
STEP 2 (if necessary)				
reprobe	TaqI[a]	U6.2	(DXS304)	0.03 distal
reprobe	XmnI	RN1	(DXS369)	0.04 proximal
STEP 3 (if necessary)				
	BanI	U6.2-20E	(DXS304)	0.03 distal
	MspI	VK21C	(DXS296)	0.02 distal
reprobe	TaqI	pc2S15	(IDS)	0.02 distal
reprobe	TaqI	RN1	(DXS369)	0.04 proximal
	HindIII	VK23B	(DXS297)	0.04 proximal

Notes

a the enzyme/probe combinations of TaqI/VK21A and TaqI/U6.2 could be replaced with MspI/VK21C and MspI/U6.2.

b distal/proximal indicates the position of the locus relative to FRAXA.

different restriction endonuclease, and using the probes VK21A, pc2S15 (IDS), and VK23B. The probability of a woman being informative for one or more of these RFLPs is 80%. In the event that a woman is not informative at these loci, the digested DNA samples may be reprobed to identify polymorphisms at DXS369 and DXS304 (Step 2). Polymorphisms would be detected in a further 14% of women. Using just three enzymes and five probes, 94% of women would be heterozygous for at least one of these polymorphisms. The probability that a woman would be heterozygous at loci which flank FRAXA is 68%. Step 3 raises the proportion of women who would be heterozygous at one or more loci to more than 98%.

In presenting this diagnostic strategy, two cautions should be noted. First, careful cytogenetic examination remains crucial to avoid inaccurate diagnosis. A common fragile site has been documented in normal men and women immediately proximal to the fragile X (Ledbetter & Ledbetter 1988; Sutherland & Baker 1990). If the two fragile sites are not distinguished, an individual may be incorrectly classified as having the fragile X syndrome or being a carrier, and subsequent genetic risk estimates based on DNA studies could be incorrect. Second, the fragile X syndrome is a complex genetic disorder. In all but the simplest of counselling situations, it is advisable to use appropriate computer programs (such as LINKAGE [Lathrop et al. 1985]) to integrate the pedigree, cytogenetic, and DNA polymorphism data to

provide accurate genetic risk estimates (Mulley et al. 1987). In view of the narrow confidence interval for FRAXA location, it is sufficient to calculate risk estimates for a single location of FRAXA, rather than at a range of FRAXA positions. This is in contrast to the situation discussed in Chapter 3.

Conclusion

The physical and genetic distances in this region of the genome were correlated in normal families in Chapter 5, and were consistent with the genome average of 1 cM per 1000 kb of DNA. If this relationship between physical and genetic distances is maintained near FRAXA, FRAXA is approximately two thousand kb proximal to VK21. The successful cloning of the gene responsible for cystic fibrosis (Rommens et al. 1989) has demonstrated that it is feasible to cover a distance such as this, and so to isolate the fragile X mutation itself.

Chapter 7.

LINKAGE HOMOGENEITY NEAR FRAXA IN NORMAL AND FRAGILE X FAMILIES.

SUMMARY

Until recently, few polymorphic loci had been genetically mapped close to FRAXA. This has been attributed to preferential recombination at Xq27, possibly in association with the fragile X mutation. The frequency of recombination around FRAXA has also been reported to vary among fragile X families. These observations suggested that the genetic map at Xq27 was different in normal versus fragile X populations, and that the genetic map also varied within the fragile X population. Such variability would complicate the calculation of carrier risk estimates on the basis of DNA studies in fragile X families.

Five polymorphic loci have now been mapped to within 4 centimorgan of FRAXA - DXS369, VK23, VK21, IDS, and DXS304. The frequency of recombination at Xq27 was evaluated using data at these loci and at more distant loci from 112 families with the fragile X syndrome. Two-point and multipoint linkage analyses failed to detect any difference in the recombination fractions in fragile X versus normal families. Two-point and multipoint tests of linkage homogeneity failed to detect any evidence of linkage heterogeneity in the fragile X families. Therefore, genetic maps derived from large samples of normal families or fragile X families are equally valid as the basis for estimating genetic risks in a particular family.

INTRODUCTION

The fragile X syndrome is characterised by unusual clinical and cytogenetic features. The clinical and cytogenetic penetrance of the mutant locus, FRAXA, is incomplete in both males and females, and varies according to the sex and intellectual status of the carrier parent (Sherman et al. 1985). FRAXA is located at - or very close to - the fragile X at Xq27.3 (Sutherland & Hecht 1985). This region of the X chromosome has been the focus of numerous genetic linkage studies in both normal and fragile X families, and it has been suggested that the frequency of recombination in this region may also be unusual.

The proposals regarding frequency of recombination at Xq27 may be summarized as follows. First, *Xq27 is a region of preferential recombination in normal and fragile X families* (Hartley et al. 1984; Szabo et al. 1984; Davies et al. 1985; Oberle et al. 1985, 1987). A 300 kilobase (kb) region of preferential recombination has been documented at Xq26 (Nguyen et al. 1989), but there is no evidence of increased recombination at distal Xq27 (Chapter 5) or at Xq28 in normal families (Feil et al. 1990). A two-point linkage study of F9 (at Xq26) and DXS52 (at Xq28) in normal and fragile X families failed to detect a difference in the recombination fraction (Oberle et al. 1986b). The recombination fraction

between these two loci is approximately 0.3, and this comparison has not been repeated using loci at Xq27.

The second proposal is that *the frequency of recombination at Xq27 varies between fragile X families.* In a series of papers, Brown et al. (1985, 1986, 1987b, 1988) documented linkage heterogeneity in the two-point recombination fraction between F9 and FRAXA. However, a multipoint analysis failed to corroborate this finding (Clayton et al. 1988). This analysis has not been repeated with polymorphic loci closer to FRAXA than F9.

A number of polymorphic loci have been localised at Xq27-q28 by physical mapping (Chapter 4) and by linkage studies in normal families (Chapter 5). The genetic map at Xq26-q28 in normal families is cen-F9-(6.2)-DXS105-(4.1)-DXS98-(14.1)-DXS369-(0.0)-VK23-(6.1)-VK21-(0.0)-IDS-(1.2)-DXS304-(13.7)-DXS52-qter (distances in centimorgan using Haldane's mapping function). FRAXA lies 2.2 centimorgan (cM) proximal to VK21 (Chapter 6).

The collaborative study reported in Chapter 6 provided a rare opportunity to examine both the frequency of recombination and linkage heterogeneity at Xq27 in a large number of fragile X families.

MATERIALS AND METHODS

The pedigree and genotype data presented in Chapter 6 and detailed in Appendix C were used for this analysis. The details of homogeneity testing are presented in Chapter 2.

RESULTS

Frequency of recombination in normal vs. fragile X families.

Two-point linkage analyses were performed using the program ILINK. The peak LOD scores and best estimates of the recombination fractions for all pairs of loci in the 112 fragile X families are summarized in Table 7-A. For comparison, the results of two-point linkage analyses of the same loci in the 40 CEPH families (Chapter 5) are also presented. There were 16 pairs of loci where the estimated recombination fractions in the two populations differed by 0.05 or more. However, the peak LOD scores were generally low (<3.00) in one or both of the populations, and little significance could be attached to the differences in recombination fractions.

One pair of loci, F9:DXS105, had a major difference in recombination fractions in the two populations, and the peak LOD scores in the two populations were both above 10.0. The program MTEST was used to determine whether this difference in recombination fraction was significant; the two recombination fractions were not significantly different (X^2_1=2.98; p=0.08).

DXS369, VK23, VK21, IDS, and DXS304 were documented to lie within 4 cM of FRAXA in Chapter 6. Forty nine of the fragile X families were informative at two or more of the polymorphic

Table 7-A

Summary of two-point linkage analysis of loci at Xq27-q28 in fragile X and CEPH families.

	F9	DXS105	DXS98	DXS369	VK23	FRAXA	VK21	IDS	DXS304	DXS52
F9		.104 .043	.041 .078	.211 .221	ND .251	.190 ND	.300 .190	.209 .193	.099 .241	.333 .327
DXS105	10.63 14.34		.045 .000	.211 .170	.155 .179	.152 ND	.350 .142	.120 .209	.306 .170	.278 .279
DXS98	5.01 6.38	3.56 6.62		.090 .000	.000 .072	.058 ND	.123 .152	.062 .125	.274 .069	.194 .237
DXS369	1.46 2.54	2.03 6.79	2.28 2.71		ND .000	.066 ND	.062 .063	.000 .054	.093 .062	.184 .124
VK23	ND 1.36	1.34 1.38	3.38 2.65	ND 4.52		.042 ND	.033 .155	.066 .077	.000 .042	.500 .175
FRAXA	6.49 ND	5.40 ND	6.70 ND	14.62 ND	4.46 ND		.015 ND	.089 ND	.031 ND	.126 ND
VK21	0.26 1.28	0.59 2.28	2.61 1.19	9.54 6.08	5.76 1.19	33.45 ND		.000 .000	.019 .000	.092 .154
IDS	1.86 4.88	1.65 4.77	2.52 4.59	0.90 15.19	3.18 6.85	4.19 ND	4.10 15.95		.000 .014	.076 .136
DXS304	0.86 2.62	0.45 5.50	0.32 5.27	3.91 12.74	0.60 5.12	9.67 ND	11.85 7.53	0.92 18.20		.161 .127
DXS52	2.62 1.46	3.16 3.55	4.40 1.81	7.00 10.83	0.00 2.36	21.45 ND	23.11 5.06	7.79 13.93	5.10 11.25	

Note: The figures above the diagonal are the best estimates of the recombination fractions between the specified loci. The figures below the diagonal are the peak LOD scores. In each case, the upper figure is the result of analysis in the fragile X families, and the lower figure is from analysis of the CEPH families. (ND no data)

loci. The recombination fractions between these five loci were estimated simultaneously in the fragile families using the program ILINK. The recombination fractions were as follows: DXS369-(0.000)-VK23-(0.069)-VK21-(0.000)-IDS-(0.018)-DXS304. The recombination fractions between these loci had been estimated in the CEPH families (Chapter 5), and were DXS369-(0.000)-VK23-(0.057)-VK21-(0.000)-IDS-(0.012)-DXS304. In the fragile X families, the difference in the transformed likelihoods (-2ln(likelihood)) of these two sets of recombination fractions was not significant (X^2_4=0.35; p>0.5).

Tests of linkage homogeneity in the fragile X families.

Tests of homogeneity were not performed for all two-point analyses in the fragile X families. These tests have limited power (Ott 1986; Risch 1988), and there is little point in performing the test with limited data. Moreover, performing multiple comparisons would have reduced the power of the tests even further. Four loci were informative in more than 40 of the fragile X families, F9, DXS369, VK21 and DXS52. Two-point homogeneity tests were performed for the recombination fractions between each of these loci and FRAXA. The data for each test were two-point LOD scores at 10 values of θ between 0.00 and 0.45. As the DXS296:FRAXA LOD score curve had a sharp peak at $\hat{\theta}$=0.015 (Table 6-B), LOD scores at θ=0.01, 0.02, 0.03, and 0.04 were also included.

The results of the homogeneity tests are summarized in Table 7-B. For each of the four loci, the recombination fraction between the locus and FRAXA was significantly less than 0.50, i.e. H_0 was rejected in favor of H_1. However, for none of the loci was there evidence of linkage heterogeneity, i.e. there was insufficient evidence to reject H_1 in favor of H_2. In Table 7-B the estimates of $\hat{\theta}$ between each locus and FRAXA under the hypothesis of linkage homogeneity (H_1) are slightly different from those listed in Table 7-A. The reason for this is that the data entry for the HOMOG programs consists of LOD scores at specified recombination fractions (calculated with the program MLINK). The estimate of $\hat{\theta}$ determined by HOMOG2 is limited to one of the specified θ values. On the other hand, the values in Table 7-A were determined iteratively using the program ILINK, and there is no restriction on the estimated value of $\hat{\theta}$.

Multipoint homogeneity testing of FRAXA location was performed with multipoint LOD scores from 101 of the fragile X families. Each of these families was informative at one or more of the loci DXS98, DXS369, VK23, VK21, IDS and DXS304. Under the hypothesis of a single location for FRAXA (H_1), the most likely location of FRAXA was 2.2 cM proximal to VK21. The odds in favor of H_1 vs. H_0 were 10^{48}:1. Under the hypothesis (H_2) of two locations for FRAXA, FRAXA was located at DXS98 in 15% of the families and 1.6 cM proximal to VK21 in the remainder. The odds in favor of H_2 vs. H_1 were 2:1.

Note to Table

The best estimates of θ and of the proportion of the families in each group were obtained under each hypothesis, and are listed in the Table. The significance of differences in the likelihood of the different hypotheses are indicated to the right of the Table.

Table 7-B

Homogeneity tests of two-point recombination fractions in the fragile X families.

		ln(relative likelihood)	
F9:FRAXA	(44 families)		
H_0:	loci unlinked in all families	0.00	
H_1:	loci linked at $\hat{\theta}$=0.20 in all families	14.96	χ^2_1=29.9 (p<0.0001)
H_2:	loci linked at $\hat{\theta}$=0.00 in 5% of families, at $\hat{\theta}$=0.20 in 95% of families.	15.04	χ^2_2=0.16 (p>0.4)
DXS369:FRAXA	(45 families)		
H_0:	loci unlinked in all families	0.00	
H_1:	loci linked at $\hat{\theta}$=0.05 in all families	33.39	χ^2_1=66.8 (p<0.0001)
H_2:	loci linked at $\hat{\theta}$=0.00 in 35% of families, at $\hat{\theta}$=0.10 in 65% of families.	33.77	χ^2_2=0.76 (p>0.3)
VK21:FRAXA	(67 families)		
H_0:	loci unlinked in all families	0.00	
H_1:	loci linked at $\hat{\theta}$=0.02 in all families	76.89	χ^2_1=153.8 (p<0.0001)
H_2:	loci linked at $\hat{\theta}$=0.01 in 48% of families, at $\hat{\theta}$=0.02 in 52% of families.	77.00	χ^2_2=0.22 (p>0.4)
DXS52:FRAXA	(89 families)		
H_0:	loci unlinked in all families	0.00	
H_1:	loci linked at $\hat{\theta}$=0.15 in all families	48.83	χ^2_1=97.7 (p<0.0001)
H_2:	loci linked at $\hat{\theta}$=0.05 in 50% of families, at $\hat{\theta}$=0.20 in 50% of families.	49.96	χ^2_2=2.27 (p>0.1)

The odds in favor of there being three locations for FRAXA rather than one were also 2:1.

DISCUSSION

The fragile X syndrome is the only clinical disorder known to associated with expression of a fragile site (Sutherland & Hecht 1985). It is also unique in being the only X-linked disorder with incomplete penetrance in males. It was therefore intriguing when linkage analysis suggested that the linkage relationships around FRAXA were unusual.

Recombination fractions in normal vs. fragile X families

Investigation of the linkage relationships around FRAXA is dependent on the availability of DNA probes which detect polymorphisms in the region. Clusters of polymorphic loci were identified at Xq26 and Xq28 (Szabo et al. 1984; Oberle et al. 1985, 1987), but until recently there have been few loci mapped close to FRAXA by linkage analysis. This lack of polymorphic loci has been attributed to either a high frequency of recombination in the region of FRAXA, or to a selection bias in isolating probes from Xq27 (Oberle et al. 1987). The region around FRAXA could contain repeated sequences, and hence be under-represented when screening for unique DNA probes from genomic libraries (Hyland et al. 1989).

A region of preferential recombination has been identified immediately proximal to F9 at Xq26 (Nguyen et al. 1989). A second region of preferential recombination was described at

Xq28 (Bell et al. 1989), but this has now been attributed to possible genotyping errors (Feil et al. 1990). In normal families, a correlation of the frequency of recombination with physical distances immediately distal to FRAXA (Chapter 5) did not detect a region of preferential recombination. In comparing normal and fragile X families, Oberle et al. (1986b) found no difference in the frequency of recombination between F9 and DXS52. This study was limited to a total of less than 30 families and involved loci that are far from FRAXA.

A comparison of two-point and multipoint linkage relationships of loci close to FRAXA in 40 normal families and over 40 fragile X families (Table 7-A) failed to detect any difference in the recombination fractions in normal versus fragile X families. Thus there is little evidence the support the contention that Xq27 is a region of preferential recombination in either normal families or in fragile X families.

Linkage heterogeneity in fragile X families

The second proposal considered in this study was whether the frequency of recombination around FRAXA varies among fragile X families. The first study of the recombination fraction between FRAXA and F9 estimated the recombination fraction to be zero (Camerino et al. 1983), but later investigations indicated that the recombination fraction was much higher

(Choo et al. 1984; Warren et al. 1985). Subsequently it was suggested that the frequency of recombination in this interval may vary among fragile X families (Brown et al. 1985, 1986, 1987b, 1988; Oberle et al. 1986b). In an analysis of 106 families, 20% had no recombination between F9 and FRAXA, while the remainder had a recombination fraction of 0.35 (Brown et al. 1988). There was no evidence of linkage heterogeneity between FRAXA and the distal locus DXS52. All of these studies were limited to analyzing loci that have recombination fractions of more than 0.12 with FRAXA.

In the present study, tests of two-point linkage homogeneity between FRAXA and F9 or DXS52 failed to document linkage heterogeneity (Table 7-B). Tests of two-point linkage homogeneity using the closely linked loci DXS369 and VK21 also failed to document heterogeneity in the fragile X families.

In the absence of a clear clinical or cytogenetic distinction among fragile X families, there is no *a priori* reason why there should be just two groups of fragile X families rather than many groups. In any linkage study, the peak LOD score will occur at a different recombination fraction in each family studied. This reflects the different pedigree structures, numbers of informative women in each family, and the stochastic nature of recombination. However, in a multipoint test of linkage homogeneity based on six loci

close to FRAXA and 101 families, the odds in favor of there being two or three locations for FRAXA rather than one were only 2:1.

The only other multipoint test of linkage homogeneity in the fragile X syndrome also failed to detect heterogeneity. Clayton et al. (1988) took essentially the same data set as Brown et al. (1987b), and performed a multipoint test of homogeneity similar to that described for the present study. The odds in favor of there being two loci for FRAXA were only 2:1. In discussing the initial conflicting reports of the F9:FRAXA recombination fraction, the authors considered the possibility that there is a familial predisposition to recombination at Xq27. Such a predisposition need not be specific to the fragile X syndrome, but may be a feature of the normal population. However, two-point tests of linkage homogeneity in the CEPH families found no evidence of heterogeneity at Xq27 (Chapter 5).

It is not clear why the conclusion of this study differs from those of other investigators. In assembling these data for analysis, there was a selective bias in favor of families informative at loci close to FRAXA. There was no bias in terms of the clinical or cytogenetic characteristics of affected males. There are three possible explanations for the discrepancy. The first is that these results represent some statistical fluctuation. The second possibility is that this

study lacked sufficient power to detect linkage heterogeneity, and is discussed below. The third possibility is that the Admixture test may not always be conservative, and that some instances of supposed linkage heterogeneity are incorrect. The possibility of obtaining a non-conservative result with the Admixture test is presented for completeness, but this seems an unlikely explanation for the linkage heterogeneity reported by the investigators who used the Admixture test (Brown et al. 1987b, 1988).

The power of a two-point homogeneity test varies according to the hypothesis being tested. Tables have been published giving the power of various tests for different hypotheses (Cavalli-Sforza & King 1986; Ott 1986; Risch 1988), but the specific hypothesis proposed by Brown (1988) is not listed. Moreover, the distribution of likelihood ratios in multipoint tests of linkage homogeneity is not known, and the power of this analysis cannot be estimated. On the other hand, this analysis was based on a large number of families. Two-point linkage heterogeneity between F9 and FRAXA has been documented in as few as six large families (Oberle et al. 1986), and the study of Brown et al. (1987b) included just 32 families.

Homogeneity tests are generally conservative (Risch 1988). However, there is one situation where the analysis may be non-conservative. This is not a problem of test design but

reflects the nature of the data entry, and has not been described elsewhere. The data required by all the HOMOG programs are LOD scores at specified values of θ for each family in the analysis. The best estimate of θ ($\hat{\theta}$) determined by HOMOG is limited to one of the specified θ values. This may differ from the estimate of $\hat{\theta}$ determined iteratively using ILINK where there is no restriction on the possible value of $\hat{\theta}$. If the two estimates of $\hat{\theta}$ differ and if the LOD curve has a sharp peak, the peak LOD score at $\hat{\theta}$ estimated by HOMOG may be lower than the peak LOD score estimated by ILINK. The peak LOD score estimated by HOMOG is transformed to be the likelihood of hypothesis H_1. If HOMOG underestimates the peak LOD score, the likelihood of H_1 will also be underestimated. The likelihood of H_2 will not be significantly underestimated as the LOD score curve for individual families is usually not sharp, and an incorrect estimate of $\hat{\theta}$ for a single family is not reflected in a major error in the peak LOD score.

The net effect of this is that the calculated difference in the likelihoods of H_1 and H_2 may be incorrectly large, and H_1 may be incorrectly rejected in favor of H_2. As an example, a two-point test of linkage homogeneity between FRAXA and VK21 rejected H_1 in favor of H_2 ($X^2_2=5.9$; $p=0.03$) if LOD scores were entered for just 10 values of θ at intervals of 0.05. If LOD scores at values of θ close to 0.015 were also included, there was insufficient evidence to reject H_1 in favor of H_2

(Table 7-B). This type of error can be avoided by including LOD scores at the recombination fraction estimated iteratively by ILINK, and by checking that the likelihood of H_1 corresponds to the peak LOD score.

It is unlikely that the linkage heterogeneity reported between F9 and FRAXA by Brown et al. in 1987(b) was due to a non-conservative result of the Admixture test: the value of $\hat{\theta}$ estimated under H_1 by the HOMOG program was the same as that determined by two-point linkage analysis ($\hat{\theta}$=0.21), and the LOD score curve was not sharply curved (the peak LOD score was 7.62). The estimate of $\hat{\theta}$ under H_1 was not presented in the paper by Brown et al. (1988), and no conclusion can be drawn regarding a possible non-conservative result. The Admixture test was used by Brown et al. in 1985, but significant heterogeneity was not detected with this test. However, heterogeneity was detected using Morton's test in a subset of large families[1].

Conclusion

For those involved in genetic counselling of families with the fragile X syndrome, the general applicability of the genetic map at Xq27 is of paramount importance. It is essential that an accurate genetic map be used when

---

1 The non-conservative effect of insufficient data for the Admixture test also applies to Morton's homogeneity test. However, the effect is masked because of the conservative nature of the test (as detailed in Chapter 2).

calculating multipoint risk estmates (Suthers & Wilson 1990 [Appendix D]). The conclusions of this analysis are encouraging. There is no evidence that the genetic map at Xq27 is different in fragile X families versus the normal population, nor is there evidence of linkage heterogeneity among fragile X families. Therefore, genetic distances that have been estimated in either normal families or fragile X families can be incorporated into genetic risk analyses with confidence.

Chapter 8

CONCLUSION.

When Lehrke presented his hypotheses regarding X-linked mental retardation (Lehrke 1972a), he indicated the biological and clinical significance of this field of study. In the area of human biology, he postulated that there are a number of genes on the X chromosome that are primarily involved in brain function, and that allelic variation of these genes could account for a significant component of the genetic basis of intelligence.

The localization of MRX1 to a region far removed from FRAXA (Chapter 3) indicates that non-specific X-linked mental retardation is genetically heterogeneous. This supports Lerhke's proposal, but the significance of these loci cannot be readily tested until the genes are cloned. Once a gene is cloned it would be possible to directly assess the impact of allelic variation at the locus, and to begin evaluating the cell biology of the gene product.

In view of the genetic heterogeneity of non-specific X-linked mental retardation, the fragile X mutation, FRAXA, is the only feasible target for a research program directed at cloning such a gene. The panel of cell lines described in Chapter 4 and the precise localization of probes near FRAXA has already enabled large scale restriction maps to be generated near FRAXA (Hyland et al. submitted). In collaboration with Dr D Schlessinger (St. Louis, USA), it has been possible to isolate yeast artificial chromosomes (YACS)

with the probes VK16, VK18, VK21, and VK23 (E Cremer, M Pritchard, S Yu; Adelaide, S.A.). It may be possible to span FRAXA with a series of overlapping YACS. The YACS and the VK probes are also the basis of linking and jumping libraries being developed around FRAXA (M Pritchard). As the large scale map is developed across FRAXA, conserved sequences can be isolated and examined in normal and fragile X families.

Lerhke also emphasized the clinical significance of the X-linked brain-specific genes. Mutations at these genes are common, and - in the absence of an unequivocal diagnostic marker - it can be difficult to offer genetic advice to consultands in the families. The mapping of genes responsible for non-specific X-linked mental retardation syndromes is of little benefit to consultands unless the information can be used in genetic risk estimation. The resampling procedure presented in Chapters 2 and 3 provides a method of obtaining reliable genetic risk estimates from very limited data, and will have application in the counselling of other rare syndromes.

The probes VK21, VK23, and pc2S15 (Chapter 5) are already in use in providing genetic risk advice to fragile X families. These probes are the closest probes to FRAXA and identify polymorphisms in over 80% of women (Chapter 6). The VK probes may also identify highly polymorphic loci of the type described by Litt and Luty (1989) and by Weber and May (1989)

(RI Richards, Adelaide, S.A.). This will increase the proportion of women heterozygous for loci very close to FRAXA, and render other probes in the area redundant. With the combination of highly informative loci close to FRAXA and a reliable genetic map at Xq27 (Chapter 7), it will be possible to provide prompt and specific genetic advice to even more men and women from fragile X families.

REFERENCES

Abruzzo MA, Pettay D, Mayer M, Jacobs PA (1986). The effect of caffeine on fragile X expression. Hum Genet 73:20-22.

Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD (1983). Molecular Biology of the Cell. New York: Garland Publishing, Inc.

Aldridge J, Kunkel L, Bruns G, et al. (1984). A strategy to reveal high-frequency RFLPs along the human X chromosome. Am J Hum Genet 36:546-564.

Anastasi A (1972). Four hypotheses with a dearth of data: response to Lehrke's 'A theory of X-linkage of major intellectual traits'. Amer J Ment Defic 76:620-622.

Arveiler B, Oberle I, Vincent A, Hofker MH, Pearson PL, Mandel JL (1988a). Genetic mapping of the Xq27-28 region: New RFLP markers useful for diagnostic applications in fragile-X and hemophilia-B families. Am J Hum Genet 42:380-389.

Arveiler B, Alembik Y, Hanauer A, et al. (1988b). Linkage analysis suggests at least two loci for X-linked non-specific mental retardation. Am J Med Genet 30:473-483.

Bakker E, Wieacker P, Beverstock GC, Pearson PL (1983). Recombinant DNA techniques for mapping the human X chromosome. Clin Genet 23:225.

Bateson W, Saunders ER, Punnett RC (1905). Experimental studies in the physiology of heredity. Reports of the Evolution Committee of the Royal Society. 2:80-99.

Bell J, Haldane JBS (1948). The linkage between the genes for colour-blindness and haemophilia in man. Proc Roy Soc Lond B 123:119-150.

Bell MV, Patterson MN, Dorkins HR, Davies KE (1989). Physical mapping of DXS134 close to the DXS52 locus. Hum Genet 82:27-30.

Benham F, Hart K, Crolla J, Bobrow M, Francavilla M, Goodfellow PN (1989). A method for generating hybrids containing non-selected fragments of human chromosomes. Genomics 4:509-517.

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

Blomquist HK, Gustavson KH, Holmgren G, Nordenson I, Sweins A (1982). Fragile site X chromosomes and X-linked mental retardation in severely retarded boys in a northern Swedish county. A prevalence study. Clin Genet 21:209-214.

Boehnke M (1986). Estimating the power of a proposed linkage study: A practical computer simulation approach. Am J Hum Genet 39:513-527.

Boggs BA, Nussbaum RL (1984). Two anonymous X-specific human sequences detecting restriction fragment length polymorphisms in region Xq26-qter. Somatic Cell Mol Genet 10:607-613.

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

Boyd Y, Buckle V, Holt S, Munro E. Hunter D, Craig I (1986). Muscular dystrophy in girls with X;autosome translocations. J Med Genet 23:484-490.

Brown CJ, Willard HF (1988). The human A1S9T gene escapes X chromosome inactivation: Regional localisation to distal Xp and use as a selectable marker in the creation of novel somatic cell hybrids. Am J Hum Genet 43(Suppl):A103.

Brown WT, Gross AC, Chan CB, Jenkins EC (1985). Genetic linkage heterogeneity in the fragile X syndrome. Hum Genet 71:11-18.

Brown WT, Gross AC, Chan CB, Jenkins EC (1986). DNA linkage studies in the fragile X syndrome suggest genetic heterogeneity. Am J Med Genet 23:643-664.

Brown WT, Wu Y, Gross AC, Chan CB, Dobkin CS, Jenkins EC (1987a). RFLP for linkage analysis of fragile X syndrome. Lancet i:280.

Brown WT, Jenkins EC, Gross AC, et al. (1987b). Further evidence for genetic heterogeneity in the fragile X syndrome. Hum Genet 75:311-321.

Brown WT, Sherman SL, Dobkin CS (1987c). Hypothesis regarding the nature of the fragile X mutation: A reply to Winter and Pembrey. Hum Genet 75:294-295.

Brown, WT, Gross A, Chan C, et al. (1988). Multilocus analysis of the fragile X syndrome. Hum Genet 78:201-205.

Brown WT (1989). DNA studies of the fragile X mutation. In: KE Davies, ed. The Fragile X Syndrome. Oxford: Oxford University Press, 76-101.

Brown WT, Gross A, Goonewardena P, Ferrando C, Dobkin C, Jenkins EC (1990). Detection of fragile X non-penetrant males by DNA marker analysis (submitted).

Buetow KH, Chakravarti A (1987). Multipoint gene mapping using seriation. II. Analysis of simulated and empirical data. Am J Hum Genet 41:189-201.

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

Camerino G, Mattei MG, Mattei JF, Jaye M, Mandel JL (1983). Close linkage of fragile X-mental retardation syndrome to haemophilia B and transmission through a normal male. Nature 306:701-704.

Cavalli-Sforza LL, King MC (1986). Detecting linkage for genetically heterogeneous diseases and detecting heterogeneity with linkage data. Am J Hum Genet 38:599-616.

Choo KH, George D, Filby G, et al. (1984). Linkage analysis of X-linked mental retardation with and without fragile-X using factor IX gene probe. Lancet ii:349.

Chudley AE, Knoll J, Gerrard JW, Shepel L, McGahey E, Anderson J (1983). Fragile (X) X-linked mental retardation I: Relationship between age and intelligence and the frequency of expression of fragile (X)(q28). 14:699-712.

Chudley AE, Hagerman RJ (1987). Fragile X syndrome. J Pediatr 110:821-830.

Clayton JR (1986). A multipoint linkage analysis program for X-linked disorders, with the example of Duchenne muscular dystrophy and seven DNA probes. Hum Genet 73:68-72.

Clayton JF, Gosden CM, Hastie ND, Evans HJ (1988). Linkage heterogeneity and the fragile X. Hum Genet 78:338-342.

Clerget-Darpoux F, Bonaiti-Pellie C, Hochez J (1986). Effects of misspecifying genetic parameters in LOD score analysis. Biometrics 42:393-399.

Cohen SN, Chang ACY, Boyer HW, Helling RB (1973). Construction of biologically functional bacterial plasmids in vitro. Proc Nat Acad Sci USA 70:3240-3244.

Conneally PM, Rivas ML (1980). Linkage analysis in man. In: Harris, Hirschhorn, eds. Advances in Human Genetics 10. New York: Plenum Press, pp. 209-266.

Conneally PM, Edwards JH, Kidd KK, et al. (1985). Report of the committee on methods of linkage analysis and reporting. Eighth International Workshop on Human Gene Mapping. Cytogenet Cell Genet 40:356-359.

Cooper DN, Schmidtke J (1984). DNA restriction fragment

length polymorphisms and heterozygosity in the human genome. Hum Genet 66:1-16.

Couturier J, Dutrillaux B, Garber P, et al. (1979). Evidence for a correlation between late replication and autosomal gene inactivation in a familial translocation t(X;21). Hum Genet 49:319-326.

Cremers FPM, van de Pol TJR, Wieringa B, et al. (1988). Molecular analysis of male-viable deletions and duplications allows ordering of 52 DNA probes on proximal Xq. Am J Hum Genet 43:452-461.

Crow JF (1988). A diamond anniversary: The first chromosome map. Genetics 118:1-3.

Dahl N, Hammarstrom-Heeroma K, van Ommen GB, Petterson U (1989a). A polymorphic locus at Xq27-28 detected by the probe U6.2 (DXS304). Nucl Acid Res 17:2884.

Dahl N, Goonewardena P, Malmgren H, et al. (1989b). Linkage analysis of families with fragile-X mental retardation, using a novel RFLP marker (DXS304). Am J Hum Genet 45:304-309.

Davidson RL, Gerald PS (1976). Improved techniques for the induction of mammalian cell hybridization by polyethylene glycol. Somat Cell Genet 2:165-176.

Davies KE, Pearson PL, Harper PS, et al. (1983). Linkage analysis of two cloned DNA sequences flanking the Duchenne muscular dystrophy locus on the short arm of the human X chromosome. Nucl Acid Res 11:2302-2312.

Davies KE (1985). Molecular genetics of the human X chromosome. J Med Genet 22:243-249.

Davies KE, Mattei MG, Mattei JF, et al. (1985). Linkage studies of X-linked mental retardation: High frequency of recombination in the telomeric region of the human X chromosome (fragile site/linkage/recombination/X chromosome). Hum Genet 70:249-255.

Davison B.C.C. (1973). Genetic studies in mental subnormality. Brit J Psychiatr, special publication No. 8.

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

Demenais F, Lathrop GM, Lalouel JM (1988). Detection of linkage between a quantitative trait and a marker locus by the LOD score method: Sample size and sampling considerations. Ann Hum Genet 52:237-246.

Devor EJ (1988). The relative efficiency of restriction enzymes: An update. Am J Hum Genet 42:179-182.

DiCiccio TJ, Romano JP (1988). A review of bootstrap confidence intervals. J Roy Stat Soc B 50:338-354.

Donis-Keller H, Green P, Helms C, et al. (1987). A genetic linkage map of the human genome. Cell 51:319-337.

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

Dronamraju KR (1987). The origins of human gene mapping: With particular reference to the contributions of JBS Haldane. Genomics 1:270-276.

D'Urso M, Mareni C, Toniolo D, et al. (1983). Regulation of glucose-6-phosphate dehydrogenase expression in CHO-human fibroblast somatic cell hybrids. Somat Cell Genet 9:429-443.

Efron B, Tibshirani R (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1:54-77.

Elder JK, Southern EM (1983). Measurement of DNA length by

gel electrophoresis II: Comparison of methods for relating mobility to fragment length. Anal Biochem 128: 227-231.

Elston RC, Stewart J (1971). A general model for the analysis of pedigree data. Hum Hered 21:523-542.

Elston RC (1989). Man bites dog? The validity of maximizing LOD scores to determine mode of inheritance. Am J Med Genet 34:487-488.

Feil R, Palmieri G, d'Urso M, Heilig R, Oberle I, Mandel JL (1990). Physical and genetic mapping of polymorphic loci in Xq28 (DXS15, DXS52, and DXS134): Analysis of a cosmid clone and a yeast artificial chromosome. Am J Hum Genet 46:720-728.

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

Filippi G, Rinaldi A, Archidiacono N, Rocchi M, Balazs I, Siniscalco M (1983). Linkage between G6PD and fragile-X syndrome. Am J Med Genet 15:113-119.

Fishburn J, Turner G, Daniel A, Brookwell R (1983). The diagnosis and frequency of X-linked conditions in a cohort of

moderately retarded males with affected brothers. Am J Med Genet 14:713-724.

Friedman JM, Howard-Peebles PN (1986). Inheritance of fragile X syndrome: An hypothesis. Am J Med Genet 23:701-713.

Froster-Iskenius U, McGillivray BC, Dill FJ, Hall JG, Herbst DS (1986). Normal male carriers in the fra(X) form of X-linked mental retardation (Martin-Bell syndrome). Am J Med Genet 23:619-631.

Fryns JP, Kleczkowska A, Wolfs I, van den Berghe (1984a). Klinefelter syndrome and two fragile X chromosomes. Clin Genet 26:445-447.

Fryns JP, Kleczkowska A, Kubien E, Petit P, van den Berge H (1984b). Inactivation pattern of the fragile X in heterozygous carriers. Hum Genet 65:400-401.

Fryns J (1986). The female and the fragile X: A study of 144 obligate female carriers. Am J Med Genet 23:157-169.

Fuscoe JC, Fenwick RG, Ledbetter DH, Caskey CT (1983). Deletion and amplification of the HGPRT locus in Chinese hamster cells. Mol Cell Biol 3:1086-1096.

Fuster C, Templado S, Miro R, Barrios L, Egozcue J (1988). Concurrence of the triple-X syndrome and expression of the fragile site Xq27.3. Hum Genet 78:293.

Giraud F, Ayme S, Mattei JF, Mattei MG (1976). Constitutional chromosome breakage. Hum Genet 34:125-136.

Glover TW, Berger C, Coyle J, Echo B (1984). DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. Hum Genet 67:136-142.

Greenberg DA (1989). Inferring mode of inheritance by comparison of LOD scores. Am J Med Genet 34:480-486.

Greenberg DA, Hodge SE (1989). Linkage analysis under 'random' and 'genetic' reduced penetrance. Genet Epidemiol 6:259-264.

Greene JR, Guarente L (1987). Subcloning. In: Berger SL, Kimmel AR, eds. Guide to Molecular Cloning Techniques. Methods in Enzymology, Vol 152. Orlando: Academic Press, pp. 512-522.

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

Hagerman RJ, van Housen K, Smith ACM, McGavran (1984). Consideration of connective tissue dysfunction in the fragile X syndrome. Am J Med Genet 17:111-121.

Hagerman R (1989). Behaviour and treatment of the fragile x syndrome. In: KE Davies, ed. The Fragile X Syndrome. Oxford: Oxford University Press, 56-75.

Haldane JBS (1948). The Formal Genetics of Man. Proc Roy Soc Lond 135:147-170.

Hanahan D (1983). Studies on transformation of Escherichia coli with plasmids. J Mol Biol 166:557-580.

Harrison CJ, Jack EM, Allen TD, Harris R (1983). The fragile X: a scanning electron microscope study. J Med Genet 20:280-285.

Hartley DA, Davies KE, Drayna D, White RL, Williamson R (1984). A cytological map of the human X chromosome-evidence for non-random recombination. Nucl Acid Res 12:5277-5285.

Harvey J, Judge C, Wiener S (1977). Familial X-linked mental retardation with an X-chromosome abnormality. J Med Genet 14:46-50.

Hecht F, Fryns JP, Vlietinck RF, van den Berghe H (1986). Genetic control over fragile X chromosome expression. Clin Genet 29:191-195.

Heilig R, Oberle I, Arveiler B, Hanauer A, Vidaud M, Mandel JL (1988). Improved DNA markers for efficient analysis of fragile X families. Am J Med Genet 30:543-550.

Herbst DS, Miller JR (1980). Nonspecific X-linked mental retardation II: The frequency in British Columbia. Am J Med Genet 7:461-469.

Herbst DS, Dunn HG, Dill FJ, Kalousek DK, Krywaniuk LW (1981). Further delineation of X-linked mental retardation. Hum Genet 58:366-372.

Hinkley DV (1988). Bootstrap methods. J Roy Stat Soc B 50:321-337.

Hirst MC, Bell MV, MacKinnon RN, et al. (1990). Mapping a cerebellar degeneration related protein and DXS304 around the fragile site. Am J Med Genet (in press).

Hoegerman SF, Rary JM (1986). Speculation on the role of transposable elements in human genetic disease with particular attention to achondroplasia and the fragile X syndrome. Am J Med Genet 23:685-699.

Hofker MH, Wapenaar MC, Goor N, Bakker E, van Ommen GJB, Pearson PL (1985). Isolation of probes detecting restriction fragment length polymorphisms from X-chromosome-specific libraries: Potential use for diagnosis of Duchenne muscular dystrophy. Hum Genet 70:148-156.

Hofker MH, Skraastad MI, Bergen AAB, et al. (1986). The X chromosome shows less genetic variation at restriction sites than the autosomes. Am J Hum Genet 39:438-451.

Hofker MH, Bergen AAB, Skraastad MI, et al. (1987). Efficient isolation of X chromosome-specific single-copy probes from a cosmid library of a human X/hamster hybrid-cell line: Mapping of new probes close to the locus for X-linked mental retardation. Am J Hum Genet 40:312-328.

Hopwood JJ (1989). Enzymes that degrade heparin and heparan sulphate. In Heparin: Clinical and Biological Properties, Clinical Applications (eds. Lane, D. & Lindahl, U.). London: Edward Arnold, pp. 191-288.

Human Gene Mapping 10 (1989): Tenth International Workshop on Human Gene Mapping. Cytogenet. Cell Genet. 51:1-1148.

Hupkes PE, van Oost BA, Perdon LF, et al. (1989). New polymorphic DNA marker (DXS369) close to the fragile site

FRAXA. Human Gene Mapping 10 (1989): Tenth International Workshop on Human Gene Mapping. Cytogenet Cell Genet 51:1016.

Hyland VJ, Fernandez KEW, Callen DF, et al. (1989). Assignment of anonymous DNA probes to specific intervals of human chromosomes 16 and X. Hum Genet 83:61-66.

Hyland VJ, Nancarrow J, Callen DF, et al. Physical linkage of four loci distal to the fragile X locus (manuscript submitted).

Israel MH (1987). Autosomal supressor gene for fragile-X: An hypothesis. Am J Med Genet 26:19-31.

Jacobs PA, Glover TW, Mayer M et al. (1980). X-linked mental retardation: A study of 7 families. Am J Med Genet 7:471-489.

Jendrisak J, Young RA, Engel JD (1987). Cloning cDNA into lambda-gt10 and lambda-gt11. In: Berger SL, Kimmel AR, eds. Guide to Molecular Cloning Techniques. Methods in Enzymology, Vol 152. Orlando: Academic Press, pp. 359-371.

Jenkins EC, Brown WT, Brooks J et al. (1986). Low frequencies of apparently fragile X chromosomes in normal control cultures: A possible explanation. Expl Cell Biol 54:40-48.

Kaiser-McCaw B, Hecht F, Cadien JD, Moore BC (1980). Fragile X-linked mental retardation Am J Med Genet 7:503-505.

Keats B, Ott J, Conneally M (1989). Report of the committee on linkage and gene order. Human gene mapping 10 (1989): Tenth International Workshop on Human Gene Mapping. Cytogenet Cell Genet 51:459-502.

Kevles DJ (1986). In the Name of Eugenics. Middlesex: Pelican.

Kidd KK, Bowcock AM, Schmidtke J, et al. (1989). Report of the DNA committee and catalogs of cloned and mapped genes and DNA polymorphisms. Human gene mapping 10 (1989): Tenth international workshop on human gene mapping. Cytogenet Cell Genet 51:622-947.

Khalifa MM, Reiss AL, Migeon BR (1990). Methylation status of genes flanking the fragile site in males with the fragile-X syndrome: A test of the imprinting hypothesis. Am J Hum Genet 46:744-753.

Knoll JH, Chudley AE, Gerrard JW (1984). Fragile(X) X-linked mental retardation. II. Frequency and replication pattern of fragile (X)(q28) in heterozygotes. Am J Hum Genet 36:640-645.

Krawczak M (1987). Genetic risk and recombination - an example of non-monotonic dependency. Hum Genet 75:189-190.

Kunz BA (1982). Genetic effects of deoxyribonucleotide pool imbalances. Environmental Mutagenesis 4:695-725.

Laird C, Jaffe E, Karpen G, Lamb M, Nelson R (1987). Fragile sites in human chromosomes as regions of late-replicating DNA. Trend Genet 3:274-281.

Laird CD, Lamb MM, Thorne JL (1990). Two progenitor cells for human oogonia inferred from pedigree data and the X-inactivation imprinting model of the fragile-X syndrome. Am J Hum Genet 46:696-719.

Lander ES, Botstein D (1986). Strategies for studying heterogeneous genetic traits in humans by using a linkage map of restriction fragment length polymorphisms. Proc Natl Acad Sci USA 83:7353-7357.

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

Lange K, Boehnke M (1982). How many polymorphic genes will it take to span the human genome? Am J Hum Genet 34:842-845.

Lange K, Kunkel L, Aldridge J, Latt SA (1985). Accurate and superaccurate gene mapping. Am J Hum Genet 37:853-867.

Lange K (1986). Approximate confidence intervals for risk prediction in genetic counseling. Am J Hum Genet 38:681-687.

Lange K, Boehnke M, Weeks DE (1987). Programs for pedigree analysis. Department of Biomathematics, University of California, Los Angeles.

Lange K, Weeks D (1989). Efficient computation of lod scores: genotype elimination, genotype redefinition, and hybrid maximum likelihood algorithms. Ann Hum Genet 53:67-83.

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

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

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

Lathrop GM, Chotai J, Ott J, Lalouel JM (1987). Tests of gene order from three-locus linkage data. Ann Hum Genet 51:235-249.

Lathrop GM, Lalouel JM (1988). Efficient computations in multilocus linkage analysis. Am J Hum Genet 42:498-505.

Lathrop GM, Nakamura Y, Cartwright P, et al. (1988). A primary genetic map of markers for human chromosome 10. Genomics 2:157-164.

Ledbetter DH, Ledbetter SD, Nussbaum RL (1986a). Implications of fragile X expression in normal males for the nature of the mutation. Nature 324:161-163.

Ledbetter DH, Airhart SD, Nussbaum RL (1986b). Somatic cell hybrid studies of fragile (X) expression in a carrier female and transmitting male. Am J Med Genet 23:429-443.

Ledbetter SA, Ledbetter DH (1988). A common fragile site at Xq27: Theoretical and practical implications. Am J Hum Genet 42:694-702.

Ledbetter SA, Schwartz CE, Davies KE, Ledbetter DH (1990). New somatic cell hybrids for physical mapping in distal Xq and the fragile X region. Am J Med Genet (in press).

Lehrke R (1972a). A theory of X-linkage of major intellectual traits. Amer J Ment Defic 76:611-619.

Lehrke R (1972b). Response to Dr Anastasi and to Drs Nance and Engel. Amer J Ment Defic 76:626-631.

Lehrke RG (1974). X-linked Mental Retardation and Verbal Disability. Birth Defects Original Article Series Vol 10 No. 1. The National Foundation, March of Dimes.

Li H, Gyllensten UB, Cui X, Saiki R, Erlich HA, Arnheim N (1988). Amplification and analysis of DNA sequences in single human sperm and diploid cells. Nature 335:414-417.

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

Loesch DZ (1986). Dermatoglyphic findings in fragile X syndrome: A causal hypothesis points to X-Y interchange. Ann Hum Genet 50:385-398.

Loesch DZ, Hay DA, Sutherland GR, Halliday J, Judge C, Webb GC (1987). Phenotypic variation in male-transmitted fragile X: Genetic inferences. Am J Med Genet 27:401-417.

Loesch DZ, Hay DA (1988). Clinical features and reproductive patterns in fragile X female heterozygotes. J Med Genet 25:407-414.

Lubs HA (1969). A marker X chromosome. Am J Hum Genet 21:231-244.

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

Mandel JL, Willard HF, Nussbaum RL, et al. (1989). Report of the committee on the genetic constitution of the X chromosome. Human Gene Mapping 10 (1989): Tenth International Workshop on Human Gene Mapping. Cytogenet Cell Genet 51:384-437.

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

Mariani T (1989). Fragile sites and statistics. Hum Genet 81:319-322.

Marmur J, Rownd R, Schildkraut CL (1963). Denaturation and renaturation of deoxyribonucleic acid. Prog Nucleic Acids Res 1:231-300.

Martin JP, Bell J (1943). A pedigree of mental defect showing sex-linkage. J Neurol Neurosurg Psych 6:154-157.

Martinez MM, Goldin LR (1989). The detection of linkage and heterogeneity in nuclear families for complex disorders: one versus two marker loci. Am J Hum Genet 44:552-559.

Mendenhall W, Scheaffer RL, Wackerly DD (1986). Mathematical Statistics with Applications. Third Edition. Boston: Duxbury.

Miller SA, Dykes DD, Pulesky HF (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucl Acid Res 16:1215.

Mitchell JA, Wray J, Michalski K (1985). Brief clinical report: Neurofibromatosis and fragile-X syndrome in the same patient. Am J Med Genet 22:571-575.

Mood AM, Graybill FA, Boes DC (1974). Introduction to the Theory of Statistics. Singapore: McGraw-Hill.

Morgan TH (1911). Random segregation versus coupling in Mendelian inheritance. Science 34:384.

Morton NE (1955). Sequential tests for the detection of linkage. Am J Hum Genet 7:277-318.

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

Morton NE, Rao DC, Lang-Brown H, MacLean CJ, Bart RD, Lew R (1977). Colchester revisited: A genetic study of mental defect. J Med Genet 14:1-9.

Morton NE, MacLean CJ, Lew R, Yee S (1986). Multipoint linkage analysis. Am J Hum Genet 38:868-883.

Morton NE, Andrews VE (1989). MAP, an expert system for multiple pairwise linkage analysis. Ann Hum Genet 53:263-269.

Mossman J, Blunt S, Stephens R, Jones EE, Pembrey M (1983). Hunter's disease in a girl: association with X;5 chromosomal translocation disrupting the Hunter gene. Arch Dis Child 58:911-915.

Mulley JC, Sutherland GR (1987). Fragile X transmission and the determination of carrier probabilities for genetic counselling. Am J Med Genet 26:987-990.

Mulley JC, Gedeon AK, Thorn KA, et al. (1987). Linkage and genetic counselling for the fragile X using DNA probes 52A, F9, DX13, and ST14. Am J Med Genet 27:435-448.

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

Mulligan LM, Phillips MA, Forster-Gibson CJ, et al. (1985). Genetic mapping of DNA segments relative to the locus for the fragile-X syndrome at Xq27.3. Am J Hum Genet 37:463-472.

Murphy PA, Ruddle FH (1985). Isolation and regional mapping of random X sequences from distal human X chromosome. Somat Cell Molec Genet 11:433-444.

Nakamura Y, Leppert M, O'Connell P, et al. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235:1616-1622.

Nance WE, Engel E (1972). One X and four hypotheses: Response to Lehrke's 'A theory of X-linkage of major intellectual traits.' Amer J Ment Defic 76:623-625.

Nathans D, Smith HO (1975). Restriction endonucleases in the analysis and restructuring of DNA molecules. Annu Rev Biochem 44:273-293.

Nguyen C, Poustka AM, Djabali M, et al. (1989). Large scale

mapping and chromosome jumping in the q27 region of the human X chromosome. Genomics 5:298-303.

Nielsen KB (1983). Diagnosis of the fragile X syndrome (Martin-Bell syndrome). Clinical findings in 27 males with the fragile site at Xq28. J Ment Defic Res 27:211-226.

Nisen P, Stamberg J, Ehrenpreis R, et al. (1986). The molecular basis of severe hemophilia B in a girl. N Engl J Med 315:1139-1142.

Nobile C, Romeo G (1988). Partial digestion with restriction enzymes of ultra-violet irradiated human genomic DNA: A method for identifying restriction site polymorphisms. Genomics 3:272-274.

Noll WW, Collins M (1987). Detection of human DNA polymorphisms with a simplified denaturing gradient gel electrophoresis technique. Proc Natl Acad Sci USA 84:3339-3343.

Nussbaum RL, Airhart SD, Ledbetter DH (1983). Expression of the fragile (X) chromosome in an interspecific somatic cell hybrid. Hum Genet 64:148-150.

Nussbaum RL, Ledbetter DH (1986). Fragile X syndrome: A unique mutation in man. Ann Rev Genet 20:109-145.

Oberle I, Drayna D, Camerino G, White R, Mandel JL (1985). The telomeric region of the human X chromosome long arm: Presence of a highly polymorphic DNA marker and analysis of recombination frequency. Proc Natl Acad Sci USA 82:2824-2828.

Oberle I, Camerino G, Kloepfer C, et al. (1986a). Characterization of a set of X-linked sequences and of a panel of somatic cell hybrids useful for the regional mapping of the human X chromosome. Hum Genet 72:43-49.

Oberle I, Heilig R, Moisan JP, et al. (1986b). Genetic analysis of the fragile-X mental retardation syndrome with two flanking polymorphic DNA markers. Proc Natl Acad Sci USA 83:1016-1020.

Oberle I, Camerino G, Wrogemann K, Arveiler B, Hanauer A, Raimondi E, Mandel JL (1987). Multipoint genetic mapping of the Xq26-q28 region in families with fragile X mental retardation and in normal families reveals tight linkage of markers in q26-27. Hum Genet 77:60-65.

Oberle I, Vincent A, Abbadi N, et al. (1990). A new polymorphism and a new chromosome breakpoint establish the physical and genetic mapping of DXS369 in the DXS98-FRAXA interval. Am J Med Genet (in press).

Ogden RC, Adams DA (1987). Electrophoresis in agarose and acrylamide gels. In: Berger SL, Kimmel AR, eds. Guide to Molecular Cloning Techniques. Methods in Enzymology, Vol 152. Orlando: Academic Press, pp. 61-87.

Ohno S (1969). Evolution of sex chromosomes in mammals. Annu Rev Genet 3:495-524.

Oostra BA, Hupkes PE, Perdon LF, et al. (1990). New polymorphic DNA marker close to the fragile site FRAXA. Genomics 6:129-132.

Orita M, Suzuki Y, Sekiya T, Hayashi K (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5:874-879.

Ott J (1974). Estimation of the recombination fraction in human pedigrees: Efficient computation of the likelihood for human linkage studies. Am J Hum Genet 26:588-597.

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

Ott J (1986). The number of families required to detect or exclude linkage heterogeneity. Am J Hum Genet 39:159-165.

Ott J, Lathrop GM (1987). Goodness-of-fit tests for locus order in three-point mapping. Genet Epidemiol 4:51-57.

Pascoe L, Morton NE (1987). The use of map functions in multipoint mapping. Am J Hum Genet 40:174-183.

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726.

Patterson M, Kenwrick S, Thibodeau, et al. (1987a). Mapping of DNA markers close to the fragile site on the human X chromosome at Xq27.3. Nucl Acid Res 15:2639-2651.

Patterson M, Schwartz C, Bell M, et al (1987b). Physical mapping studies on the human X chromosome in the region Xq27-Xqter. Genomics 1:297-306.

Patterson MN, Bell MV, Bloomfield J, et al. (1989). Genetic and physical mapping of a novel region close to the fragile X site on the human X chromosome. Genomics 4:570-578.

Pembrey ME, Winter RM, Davies KE (1985). A premutation that generates a defect at crossing over explains the inheritance of fragile X mental retardation. Am J Med Genet 21:709-717.

Penrose LS (1938). A clinical and genetic study of 1280 cases of mental defect. Medical Research Council Special Report Series, Number 229, London.

Rekila AM, Vaisanen ML, Kahkonen M, Leisti J, Winqvist R (1988). A new RFLP with StuI and probe cX55.7 (DXS105) and its usefulness in carrier analysis of fragile X syndrome. Hum Genet 80:193.

Riddell DC, Wang HS, Beckett J, et al. (1986). Regional localisation of 18 human X-linked DNA sequences. Cytogenet Cell Genet 42:123-128.

Rigby PWJ, Dieckmann M, Rhodes C, Berg P (1977). Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J Mol Biol 113:237-251.

Risch N (1988). A new statistical test for linkage heterogeneity. Am J Hum Genet 42:353-364.

Roberts SH, Upadhyaya M, Sarfarazi M, Harper PS (1989). Further evidence localising the gene for Hunter's syndrome to the distal region of the X chromosome long arm. J Med Genet 26:309-313.

Rocchi M, Romeo G, Giambarrasi I, et al. (1989). Physical dissection of the human X chromosome, with special reference to the region Xq27->qter. Human Gene Mapping 10 (1989): Tenth International Workshop on Human Gene Mapping. Cytogenet Cell Genet 51:1067.

Rocchi M, Archidiacono N, Rinaldi A, et al. (1990). Mental retardation in heterozygotes for the fragile X mutation: Evidence in favour of an X inactivation-dependent effect. Am J Hum Genet 46:738-743.

Rommens JM, Ianuzzi MC, Kerem BS, et al. (1989). Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 245:1059-1065.

Rosenstraus M, Chasin LA (1975). Isolation of mammalian cell mutants deficient in glucose-6-phosphate dehydrogenase activity: Linkage to hypoxanthine phosphoribosyl transferase. Proc Natl Acad Sci USA 72:493-497.

Rosanoff AJ (1931). Sex-linked inheritance in mental deficiency. Amer J Psychiatr 11:289-297.

Rousseau F, Vincent A, Oberle I, Mandel JL (1990). New informative polymorphism at the DXS304 locus, a close distal marker for the fragile X locus. Hum Genet 84:263-266.

Schaap T (1989). The role of recombination in the evolvement of the fragile X mutation. Hum Genet 82:79-81.

Schmickel RD (1986). Contiguous gene syndromes: A component of recognizable syndromes. J Pediatr 109:231-241.

Schmidt M, Certoma A, Du Sart D, et al. (1990). Unusual X chromosome inactivation in a mentally retarded girl with an interstitial deletion Xq27. Implications for the fragile X syndrome. Hum Genet 84:347-353.

Schnur RE, Ledbetter SA, Ledbetter DH, Merry DE, Nussbaum RL (1989). New polymorphisms at the DXS98 locus and confirmation of its location proximal to FRAXA by in situ hybridization. Am J Hum Genet 44:248-254.

Schonk D, Coerwinkel-Driessen M, van Dalen I, et al. (1989). Definition of subchromosomal intervals around the myotonic dystrophy gene region at 19q. Genomics 4:384-396.

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

Sherman SL, Morton NE, Jacobs PA, Turner G (1984). The marker (X) syndrome: a cytogenetic and genetic analysis. Ann Hum Genet 48:21-37.

Sherman SL, Jacobs PA, Morton NE, et al. (1985). Further segregation analysis of the fragile X syndrome with special reference to transmitting males. Hum Genet 69:289-299.

Sherman SL, Sutherland GR (1986). Segregation analysis of rare autosomal fragile sites. Hum Genet 72:123-128.

Sherman SL (1987). A new genetic model for the fragile X syndrome involving an autosomal suppressor gene - comments on the paper by M.H. Israel. Am J Med Genet 26:33-36.

Sherman SL, Turner G, Sheffield L, Laing S, Robinson H (1988a). Investigation of the twinning rate in families with the fragile X syndrome. Am J Med Genet 30:625-631.

Sherman SL, Rogatko A, Turner G (1988b). Recurrence risks for relatives in families with an isolated case of the fragile X syndrome. Am J Med Genet 31:753-765.

Simmers RN, Smith J, Shannon MF, et al. (1988). Localization of the human G-CSF gene to the region of a breakpoint in the translocation typical of acute promyelocytic leukemia. Hum Genet 78:134-136.

Simpson NE, Newman BJ, Partington MW (1984). Fragile X

syndrome III: Dermatoglyphic studies in males, Am J Med Genet 17:195-207.

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

Soberon X, Covarrubias L, Bolivar F (1980). Construction and characterization of new cloning vehicles IV: Deletion derivatives of pBR322 and pBR325. Gene 9:287-305.

Soudek D, Partington MW, Lawson JS (1984). The fragile X syndrome I: Familial variation in the proportion of lymphocytes with the fragile site in males. Am J Med Genet 17:241-252.

Southern EM (1979). Measurement of DNA length by gel electrophoresis. Anal Biochem 100:319-323.

Steinbach P, Barbi G, Baur S, Wiedenmann A (1983a). Expression of the fragile site Xq27 in fibroblasts: I. Detection of frax(X)(q27) in fibroblast clones from males with X-linked mental retardation. Hum Genet 63:404-405.

Steinbach P, Barbi G, Baur S, Vogel W (1983b). Expression of the fragile site Xq27 in fibroblasts: II. Evidence for negative and positive clones from heterozygous females and

possible relationship between frequency and phenotype. Hum Genet 64:279-282.

Steinbach P (1986). Mental impairment in Martin-Bell syndrome is probably determined by interaction of several genes: Simple explanation of phenotypic differences between unaffected and affected males with the same X chromosome. Hum Genet 72:248-252.

Sturtevant AH (1913). The linear arrangement of six sex-linked factors in drosophila, as shown by their mode of association. J Exptl Zool 14:43-59.

Sutherland GR (1977a). Marker X chromosomes and mental retardation. N Engl J Med 296:1415.

Sutherland GR (1977b). Fragile sites on human chromosomes: Demonstration of their dependence on the type of tissue culture medium. Science 197:265-266.

Sutherland GR, Hecht F (1985). Fragile Sites on Human Chromosomes. New York: Oxford University Press.

Sutherland GR, Baker E, Fratini A (1985). Excess thymidine induces folate sensitive fragile sites. Am J Med Genet 22:433-443.

Sutherland GR, Baker E (1986a). Effects of nucleotides on expression of the folate sensitive fragile sites. Am J Med Genet 23:409-417.

Sutherland GR, Baker E (1986b). Induction of fragile sites in fibroblasts. Am J Hum Genet 38:573-575.

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

Sutherland GR (1990). The detection of fragile sites on human chromosomes. In: Advanced Techniques in Chromosome Research (ed., Adolph KW). New York: Marcel Dekker, (in press).

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

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

Suthers GK, Turner G, Mulley JC (1988a). Case report: Fragile X syndrome and nephrogenic diabetes insipidus. Am J Med Genet 30:231-236.

Suthers GK, Turner G, Mulley JC (1988b). A non-syndromal form

of X-linked mental retardation (XLMR) is linked to DXS14. Am J Med Genet 30:485-491.

Suthers GK, Hyland VJ, Baker E, Fernandez KEW, Callen DF, Sutherland GR (1988c). TaqI RFLP identified by probe VK17A (DXS294) at Xq26. Nucl Acid Res 16:11389.

Suthers GK, Callen DF, Hyland VJ, et al (1989a). A new DNA marker tightly linked to the fragile X locus (FRAXA). Science 246:1298-1300.

Suthers GK, Davies KE, Baker E, Sutherland GR (1989b). TaqI RFLP identified by probe 1A1 (DXS374). Nucl Acid Res 17:8901.

Suthers GK, Hyland VJ, Callen DF, et al. (1990a). Physical mapping of new DNA probes near the fragile X (FRAXA) with a panel of cell lines. Am J Hum Genet (in press).

Suthers GK, Oberle I, Nancarrow J, et al. (1990b). Genetic mapping of new RFLPs at Xq27-q28. (submitted)

Suthers GK, Mulley JC, Voelckel MA, et al. (1900c). Genetic mapping of new DNA probes at Xq27 defines a strategy for DNA studies in the fragile X syndrome. (submitted).

Suthers GK, Sutherland GR (1990). Recombination and the fragile X. Hum Genet (in press).

Sved JA, Laird CD (1990). Population genetic consequences of the fragile X syndrome based on the X-inactivation imprinting model. Am J Hum Genet 46:443-451.

Szabo P, Purrello M, Rocchi M, et al. (1984). Cytological mapping of the human glucose-6-phosphate dehydrogenase gene distal to the fragile-X site suggests a high rate of meiotic recombination across this site. Proc Natl Acad Sci USA 81:7855-7859.

Thompson EA, Deeb S, Walker D, Motulsky AG (1988). The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. Am J Hum Genet 42:113-124.

Turner G, Turner B, Collins E (1971). X-linked mental retardation without physical abnormality: Renpenning's syndrome. Develop Med Child Neurol 13:71-78.

Turner G, Engisch B, Lindsay DG, Turner B (1972). X-linked mental retardation without physical abnormality (Renpenning's syndrome) in sibs in an institution. J Med Genet 9:324-330.

Turner G, Turner B (1974). X-linked mental retardation. J Med Genet 11:109-113.

Turner G, Eastman C, Casey J, McLeay A, Procopis P, Turner B (1975). X-linked mental retardation associated with macro-orchidism. J Med Genet 12:367-371.

Turner G, Till R, Daniel A (1978). Marker X chromosomes, mental retardation, and macro-orchidism. N Engl J Med 299:1472.

Turner G, Brookwell R, Daniel A, Selikowitz M, Zilibowitz M (1980a). Heterozygous expression of X-linked mental retardation and X-chromosome marker fra(X)(q27). N Engl J Med 303:662-664.

Turner G, Daniel A, Frost M (1980b). X-linked mental retardation, macro-orchidism, and the xq27 fragile site. J Pediatr 96:837-841.

Turner G (1983). Historical overview of X-linked mental retardation. In: Hagerman R, McBogg P, eds. The Fragile X Syndrome. Colorado: Spectra Publishing Company, pp 1-16.

Turner G, Jacobs P (1984). Marker (X)-linked mental retardation. Adv Hum Genet 13:83-112.

Turner G, Robinson H, Laing S, Purvis-Smith S (1986). Preventive screening for the fragile X syndrome. N Engl J Med 315:607-609.

Uchida IA, Joyce EM (1982). Activity of the fragile X in heterozygous carriers. Am J Hum Genet 34:286-293.

Uchida IA, Freman VCP, Jamro H, Partington MW, Soltan HC (1983). Additional evidence for fragile X activity in heterozygous carriers. Am J Hum Genet 35:861-868.

Van Dyke DL, Weiss L (1986). Maternal effect on intelligence in fragile X males and females. Am J Med Genet 23:723-737.

Veenema H, Geraedts JPM, Beverstock GC, Pearson PL (1987a). The fragile X syndrome in a large family I: Cytogenetic and clinical investigations. J Med Genet 24:23-31.

Veenema H, Veenema T, Geraedts JPM (1987b). The fragile X syndrome in a large family II: Psychological investigations. J Med Genet 24:32-38.

Vergnaud G (1989). Polymers of random short oligonucleotides detect polymorphic loci in the human genome. Nucl Acid Res 17: 7623-7630.

Vieira J, Messing J (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259-268.

Vincent A, Kretz C, Oberle I, Mandel JL (1989a). A new polymorphic marker very closely linked to DXS52 in the q28 region of the human X chromosome. Hum Genet 82:85-86.

Vincent A, Dahl N, Oberle I, et al. (1989b). The polymorphic marker DXS304 is within 5 centimorgans of the fragile X locus. Genomics 5:797-801.

Voelckel MA, Philip N, Piquet C, et al. (1989). Study of a family with a fragile site of the X chromosome at Xq27-28 without mental retardation. Hum Genet 81:353-357.

Vogel F (1984). Mutation and selection in the marker (X) syndrome: A hypothesis. Ann Hum Genet 48:327-332.

Vogel F, Kruger J, Nielsen KB, et al. (1985). Recurrent mutation pressure does not explain the prevalence of the marker(X) syndrome. Hum Genet 71:1-6.

Vogelstein B, Gillespie D (1977). Preparative and analytical purification of DNA from agarose. Proc Natl Acad Sci USA 76:615-619.

Warburton D, Gersen S, Yu MT, Jackson C, Handelin B, Housman D (1990). Mono-chromosomal rodent-human hybrids from microcell fusion of human lymphoblastoid cells containing an inserted dominant selectable marker. Genomics 6:358-366.

Warren ST, Glover TW, Davidson RL, Jagadeeswaran P (1985). Linkage and recombination between fragile X-linked mental retardation and the factor IX gene. Hum Genet 69:44-46.

Webb TP, Bundey S, Thake A, Todd J (1986a). The frequency of the fragile X chromosome among school children in Coventry. J Med Genet 23:396-399.

Webb GC, Rogers JG, Pitt DB, Halliday J, Theobald T (1986b). Transmission of fragile (X)(q27) site from a male. Lancet ii:1231-1232.

Webb TP, Rodeck CH, Nicolaides KH, Gosden CM (1987). Prenatal diagnosis of the fragile X syndrome using fetal blood and amniotic fluid. Prenat Diagn 7:203-214.

Webb TP (1989). The epidemiology of the fragile X syndrome. In: KE Davies, ed. The Fragile X Syndrome. Oxford: Oxford University Press, 40-55.

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

White R, Leppert M, Bishop DT, et al. (1985). Construction of

linkage maps with DNA markers for human chromosomes. Nature 313:101-105.

White R, Lalouel JM (1986). Investigation of genetic linkage in human families. Adv Hum Genet 16:121-228.

White R, Lalouel JM (1988). Sets of linked genetic markers for human chromosomes. Annu Rev Genet 22:259-279.

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

Wieacker P, Davies KE, Cooke HJ, et al. (1984). Toward a complete linkage map of the human X chromosome: Regional assignment of 16 cloned single-copy DNA sequences employing a panel of somatic cell hybrids. Am J Hum Genet 36:265-276.

Wilson SR, La Scala B (1989). Multipoint linkage: Statistical evaluation. In: Elston RC, Spence MA, Hodge SE, MacClure JW, eds. Multipoint Mapping and Linkage based upon Affected Pedigree Members: Genetic Analysis Workshop 6. Progress in Clinical and Biological Research. New York: Alan R Liss.

Wilson PJ, Suthers GK, Callen DF, et al. (1990). Frequent deletions at Xq28 indicate genetic heterogeneity in Hunter syndrome. (submitted).

Wilson PJ, Morris CP, Anson DS, et al. Hunter syndrome: isolation of an iduronate-2-sulphatase cDNA clone. (submitted).

Winter RM, Pembrey ME (1986). Analysis of linkage relationships between genetic markers around the fragile X locus with special reference to the daughters of normal transmitting males. Hum Genet 74:93-97.

Winter (1987). Population genetics implications of the premutation hypothesis for the generation of the fragile X mental retardation gene. Hum Genet 75:269-271.

Winter RM, Pembrey M (1987). Interpretation of the heterogeneity in the linkage relationships of DNA markers around the fragile.X locus. Hum Genet 77:297-298.

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

Yu S, Suthers GK, Mulley JCM (1990). A BclI RFLP for DXS296 (VK21) near the fragile X. Nucl Acid Res 18:690.