Comparative PCR analysis of two triplet repeat loci and factors influencing their mutability

A thesis submitted for the degree of Master of Science of the University of Glasgow

by

Fiona Kathryn Gould B.Sc.

Division of Molecular Genetics

Institute of Biomedical and Life Sciences

February 2000

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

Fourteen expanded trinucleotide loci have been found to cause neurological diseases including Huntington's Disease, myotonic dystrophy type 1 and spinocerebellar ataxia. The repeat sequence and allele sizes vary between each locus, and the flanking DNA is assumed to play a significant role in repeat instability. The aim of this project was to characterise the behaviour and mutability of two trinucleotide repeat loci with the same repeat sequence, but with different flanking DNA sequences. Comparing the mutation rates and average length changes of these loci in the male germline demonstrated a clear effect of the influence of the flanking sequence. The *ERDA1* locus has low %GC content in the flanking sequence and displays a low mutation rate, even for large length alleles, whereas *CTG18.1* has a relatively high %GC content and demonstrates an increased mutation rate and average length change for similarly sized alleles. Utilising the sensitive SP-PCR technique gives a clear insight into how the *CTG18.1* locus has responded dramatically to defects in mismatch repair, both in MMR deficient cell lines and HNPCC tissues, and possibly in an ovarian tumour. This investigation has also revealed a possible dominant negative mutation in the *hMLH1* gene which maintains its influence on an expanded *CTG18.1* allele in two generations of a HNPCC family. In conclusion, we have successfully characterised the behaviour of two trinucleotide repeat loci which highlights the significant effect of flanking DNA sequence on the stability of triplet repeats.

Table of contents

List of Figures

List of Tables

Acknowledgements

I'd like to thank Dr. Darren Monckton for giving me the opportunity to gain a MSc and for all the support and understanding in recent years. The enthusiasm and dedication to your work encouraged me to love triplet repeats and SP-PCR! I'd also like to thank my assessors, Dr. Kevin O'Dell and Dr. Martin Boocock for their impartial opinions and for putting me at ease during assessments.

To John MacAbney who has been invaluable to me during this MSc project. John undertook the majority of the *ERDA1* sperm SP-PCR work and made many GeneScan gels. I will never forget my first student and how easy you made the initiation. Thanks also to Susan Mairs, my first official student who discovered the instability of OT481 and typed the ovarian tumours and the chimps.

To Dr. Mena Brown at the Infertility Clinic, Western Infirmary, Glasgow for the semen samples, Dr. Robert Brown at the CRC Beatson Laboratories and Dr. Margot Thomas at the Department of Obstetrics and Gynaecology, both at the Univerisity of Glasgow for supplying the ovarian tumour and cell line DNA samples, Dr. Michael Siciliano at the U.T.M.D. Anderson Cancer Center, Houston and Prof Patricia Jacobs at the Wessex Regional Genetics Laboratory, Salisbury for the HNPCC blood and tumour DNA, Dr. Tetsuo Ashizawa, Baylor College of Medicine, Houston for the DM-1 blood DNA and Dr. Marvin Meistrich at the U.T.M.D. Anderson Cancer Center, Houston for the radiotherapy sperm samples.

To Andrea Mitchell who has been a huge source of valuable experience, especially with GeneScan, and whose understanding has meant everything. To Dr. Mark Bailey, without whom GeneScan would be a mystery and for the CEPH DNA. To Heather Johnston for the advice on fixed tissue, and Graham Hamilton who has been a huge help with any computer problems that kept constantly arising. Finally, I could not express how much Simon has helped during the course of this MSc. Thanks for all the support.

The research reported in this thesis is my own original work, except where otherwise stated, and has not been submitted for any other degree.

Fiona K. Gould

February 2000

Chapter 1

Introduction

1.1 Triplet Repeats

Short tandem repeats (STRs) represent a common class of eukaryotic repetitive DNA which can range from 1-6 bp and vary in their sequence. Mononucleotide repeats are frequently made up of adenosine and dinucleotide repeats used for linkage analysis are often $(CA)_n$. Trinucleotide repeats have become the focus of much attention over the last ten years as an increasing number of neurological diseases are attributed to expansions of this type of repeat from the normal range for that locus. These trinucleotides can differ in their sequence, the purity of this sequence, their location in respect to a gene, the range of alleles detected and repeat behaviour in somatic and germline tissues. Fourteen expandable trinucleotide repeat diseases have been identified in humans so far, known as *FRAXA* (Fu *et al.*, 1991), *SBMA* (La Spada *et al.*, 1991), *DM-1* (Fu *et al.*, 1991, Aslanidis *et al.*, 1992, Brook *et al.*, 1992, Buxton *et al.*, 1992, Harley *et al.*, 1992, Mahadevan *et al.*, 1992), *FRAXE* (Flynn *et al.*, 1993), *HD* (The Huntington's Disease Collaborative Research Group, 1993), *SCA1* (Orr *et al.*, 1993), *DRPLA* (Koide *et al.*, 1994), *MJD/SCA3* (Kawaguchi *et al.*, 1994), Friedreich's ataxia (Campuzano *et al.*, 1996), *SCA2* (Imbert *et al.*, 1996, Pulst *et al.*, 1996, Sanpei *et al.*, 1996), *SCA6* (Zhuchenko *et al.*, 1997), *SCA7* (David *et al.*, 1997), *SCA8* (Koob *et al.*, 1999) and *SCA12* (Holmes *et al.*, 1999). Expandable trinucleotides at the *FRA16A* (Nancarrow *et al.*, 1994), *FRAXF* (Parrish *et al.*, 1994) and *FRA11B* (Jones *et al.*, 1994) loci do not directly cause diseases but in the case of *FRA11B* can lead to a chromosome deletion involving 11q23 that results in Jacobsen syndrome. *GDNF* (Grimm *et al.*, 1998), *PABP2* (Brais *et al.*, 1998) and *TBP* (Imbert *et al.*, 1994) have been investigated in patients with neurological disorders and although expanded alleles of 7-13 repeats at *PABP2* (Brais *et al.*, 1998) and 60 repeats at *TBP* (Koide *et al.*, 1999) have been detected, no direct association has been found with disease. Other loci such as *CTG18.1* (Breschel *et al.*, 1997), *ERDA1* (Nakamoto *et al.*, 1997, Ikeuchi et al., 1998), and *CAGR1* (Margolis *et al.*, 1996) also have expandable repeats which are not associated with a disease. Diseases such as autosomal dominant pure spastic paraplegia (Nielsen *et al.*, 1997, Benson et al., 1998), bipolar affective disorder and schizophrenia (Lindblad *et al.*, 1995, Morris *et al.*, 1995, Vincent *et al.*, 1996, Oruc *et al.*, 1997, Burgess *et al.*, 1998, Vincent *et al.*, 1998), are also hypothesised to be caused by as yet unidentified expanded trinucleotide repeat sequences.

Genes containing trinucleotide repeats are present in normal individuals, with the number of repeats ranging from several to approximately 40. The presence of an expanded allele with a trinucleotide repeat size exceeding the normal range leads to expression of disease phenotypes (Table 1.1). Anticipation is a feature of triplet repeat diseases, whereby repeat size increases in transmission are associated with an earlier onset of the disease accompanied with more severe symptoms. The repeat can be in the 5' UTR, 3' UTR, coding or non-coding regions of a gene (Figure 1.1) and have the potential to expand to greater than 1,000 repeats at loci such as *DM-1, CTG18.1, FRAXA, FRAXF, FRA11B, FRA16A* and *Friedreich's ataxia.* Hence alleles such as the *DM-1* repeat can be characterised as either normal (<40 repeats), premutation (40-50 repeats), protomutation (50-80 repeats), adult onset (>100 repeats) or congenital (>700 repeats).

1.2 Mosaicism of triplet repeats

1.2.1 Somatic Mosaicism

Somatic mosaicism is defined as any observed differences in the cells of the same tissue or the tissues of the same individual. Somatic mosaicism became an issue after observations of differences in repeat sizes of trinucleotide repeats between tissues (Anvret *et al.*, 1993, Jansen *et al.*, 1994, Thornton *et al.*, 1994, Massari *et al.*, 1995, Kinoshita *et al.*, 1996) and an increasing heterogenous smear in blood over time at the *DM-1* locus (Wong *et al.*, 1995). PCR or Southern blots of the expanded repeats demonstrated larger alleles in muscle (Anvret *et al.*, 1993, Thornton *et al.*, 1994, Massari *et al.*, 1995, Kinoshita *et al.*, 1996), brain (Thornton *et al.*, 1994), testes (Thornton *et al.*, 1994), liver (Thornton *et al.*, 1994, Kinoshita *et al.*, 1996), lung (Kinoshita *et al.*, 1996), pancreas (Kinoshita *et al.*, 1996), kidney (Kinoshita *et al.*, 1996), heart (Kinoshita *et al.*, 1996) and sperm (Massari *et al.*, 1995) compared to leukocytes. However there were no size increases in lymph node, spleen and thymus (Kinoshita *et al.*, 1996) and an apparent reduction in repeat length in the cerebellum, although this could be the progenitor allele length and hence all other tissues have expansions (Kinoshita *et al.*, 1996). It also became apparent that other triplet repeats demonstrated somatic mosaicism, though not to the same extent as *DM-1. DRPLA*, (Ito *et al.*, 1998, Tanaka *et al.*, 1999), *HD* (Telenius *et al.*, 1994, De Rooij *et al.*, 1995), *SCA1* (Cancel *et al.*, 1998) and *MJD/SCA3* (Cancel *et al.*, 1998, Ito *et al.*, 1998), also have smaller repeats in the cerebellum though no somatic mosaicism could be found in different brain regions of *SCA6* patients (Ishikawa *et al.*, 1999). *SCA1* and *FRAXA* alleles sized from blood contract with increasing progenitor repeat size (Mornet *et al.*, 1996, Koefoed *et al.*, 1998, Nolin *et al.*, 1999), whereas *SCA7* alleles have increasing numbers of extra bands with increasing allele size (Gouw *et al.*,

1998), and longer *MJD/SCA3* alleles have been detected in liver and colon (Ito *et al.*, 1998). Analysis with large quantities of DNA indicated no somatic mosaicism in the brains of *HD* patients (MacDonald *et al.*, 1993, Zuhlke *et al.*, 1993), although extra bands were detected in other studies (Telenius *et al.*, 1994, De Rooij *et al.*, 1995) and differences have now been detected using fluorescent PCR in the caudate nucleus, frontal cortex and the white matter, compared to leukocytes (Tanaka *et al.*, 1999).

Using a technique known as Small Pool PCR (SP-PCR), it is now possible to understand why larger smears are evident in the muscle of *DM-1* patients over time (Wong *et al.*, 1995) or how larger repeat bands are consistently observed in testes or sperm, irrelevant of the disease (MacDonald *et al.*, 1993, Telenius *et al.*, 1994, Massari *et al.*, 1995, Kinoshita *et al.*, 1996, Takiyama *et al.*, 1997, De Michele *et al.*, 1998, Tanaka *et al.*, 1999). Multiple aliquots of a known number of molecules of DNA undergo PCR across the repeat of interest so that the smear observed in tissues can be resolved into the component allele population (Monckton *et al.*, 1995). Using SP-PCR it has become clear that there is a tissue-specific, age and progenitor allele size effect on the repeat behaviour. SP-PCR on the *DM-1* and *SCA7* repeats demonstrates that there is a defined lower boundary of repeat contraction in muscle, blood and buccal cells which is presumed to be the progenitor allele (Monckton *et al.*, 1995, Martorell *et al.*, 1998, Monckton *et al.*, 1999). Over time, the repeat is heavily biased toward expansions which explains the larger smear detected with large quantities of DNA in muscle. SP-PCR showed that a *FRAXA* 55 repeat is relatively stable in blood (Mornet *et al.*, 1996) and single molecule PCR confirmed an increasing contraction rate with increasing allele size (Nolin *et al.*, 1999).

1.2.2 Germline Mosaicism

Anticipation within disease families suggests that triplet repeats are also unstable in the germline. There is a clear parental effect on the transmission of the *DM-1, DRPLA, SCA7* and *SCA8* repeats and the age of the transmitting parent at conception can also contribute to an increased progenitor allele size. To directly observe male germline instability of repeats, single sperm analysis can be approached in two ways. Single sperm can be physically separated, lysed and PCR (either fluorescent or radioactive) performed on the haploid genomic DNA, or DNA extracted from a sperm sample can be diluted to effectively single sperm level and PCR performed, with or without fluorescently labelled primers. Fluorescent PCR products can be separated on polyacrylamide sequencing gels and analysed by ALF Fragment Manager or GeneScan software to compare the mutation rates of single sperm to somatic DNA (frequently blood DNA) from the same individual. Normal (<50 repeats) and

expanded (>50 repeats) alleles have been analysed from affected patients at the *DRPLA* (Takiyama *et al.*, 1999), *FRAXA* (Kunst *et al.*, 1997, Nolin *et al.*, 1999), *HD* (Leeflang *et al.*, 1995, Chong *et al.*, 1997, Leeflang *et al.*, 1999), *MJD/SCA3* (Takiyama *et al.*, 1997), *SBMA* (Zhang *et al.*, 1994), and *SCA1* (Koefoed *et al.*, 1998) loci. There is a large amount of variability in the male germline between the different diseases and given the similarity in sequence, the flanking DNA is presumed to play a large part in instability (Brock *et al.*, 1999). Limited single molecule data is available for the majority of the triplet diseases, the most comprehensively covered being *HD* (Leeflang *et al.*, 1995, Chong *et al.*, 1997, Leeflang *et al.*, 1999), followed by *MJD/SCA3* (Takiyama *et al.*, 1997). The distribution of the expanded alleles broadens with increasing allele size (Leeflang *et al.*, 1999), with an increasing frequency of expansions relative to contractions in *HD* (Leeflang *et al.*, 1995, Chong *et al.*, 1997, Leeflang *et al.*, 1999), and the opposite effect is seen in *MJD/SCA3* (Takiyama *et al.*, 1997). *HD* alleles can reach a 90% mutation rate by 38-40 repeats (Leeflang *et al.*, 1995, Leeflang *et al.*, 1999) and although there is a lack of alleles to compare with in this range, 50 *SCA1* repeats, 60-62 *DRPLA* repeats, 74-75 *MJD/SCA3* repeats and 100 *FRAXA* repeats have mutation rates of 75% (Koefoed *et al.*, 1998), 96-98% (Takiyama *et al.*, 1999), 88-91% (Takiyama *et al.*, 1997) and 55% respectively, although interruptions are frequent in the *FRAXA* repeat which decreases the mutation rates of equivalently sized alleles (Nolin *et al.*, 1999). Dramatic reversions have been detected in the male germline, both in *DM-1* (Monckton *et al.*, 1995) and *Friedreich's ataxia* (De Michele *et al.*, 1998).

Therefore single molecule analysis can confirm observations in disease pedigrees, although it is not presently feasible to detect repeat behaviour in the female germline. Variations are observed between alleles of the same size from different individuals (Zhang *et al.*, 1994, Leeflang *et al.*, 1995, Chong *et al.*, 1997, Leeflang *et al.*, 1999) and though some can be attributed to age of the individual or the transmitting parent (De Michele *et al.*, 1998), others cannot (Leeflang *et al.*, 1999). Unequal transmission of the expanded allele versus the normal allele has also been seen in *SCA1* and *MJD/SCA3* disease pedigrees, resulting in the phenomenon of segregation distortion in the offspring (Riess *et al.*, 1997). The mechanism of meiotic drive has been hypothesised to contribute to segregation distortion and evidence from *DRPLA* and *MJD/SCA3* suggests this may be a factor in the male germline (Ikeuchi *et al.*, 1996) although no evidence was found for *FRAXA* or *FRAXE* in either sex (Teague *et al.*, 1998). Detailed work concerning meiotic drive and segregation distortion has been carried out at the *DM-1* locus with conflicting results. Early work suggesting segregation distortion of alleles >19 repeats from the male germline in normal pedigrees (Carey *et al.*, 1994, Gennarelli *et al.*, 1994) was reanalysed and no significant evidence for meiotic drive in males was found (Hurst *et al.*, 1995). Single sperm analysis from three heterozygotes with alleles <

and > 19 repeats found no difference in the number of smaller versus larger alleles (Leeflang *et al.*, 1996).

1.3 Locus *CTG18.1*

Locus *CTG18.1* was first identified in 1997 while searching for a candidate for Bipolar Affective Disorder (BPAD). Due to reports of anticipation and linkage to chromosome 18, a human genomic chromosome 18 library was screened with a $(CAG)_{15}$ repeat probe and a 1.6 kbp insert was identified containing a $(CTG)_{24}$ flanked immediately 3' by a non-polymorphic $(CTC)_{13}$ repeat (Breschel *et al.*, 1997). PCR was carried out with primers flanking the repeat on 56 unrelated individuals from 15 Centre d'Etude du Polymorphisme Humain (CEPH) pedigrees and 48 BPAD families which contained at least one treated BPAD patient and at least two affected first degree relatives. A subset of the BPAD families showed linkage to chromosome 18q21. *CTG18.1* demonstrated a heterozygosity of 84% with highly polymorphic allele lengths. The *CTG18.1* locus was mapped to 18q21.1 by two-point linkage analysis in 48 families ascertained for BPAD. When comparing allele sizes between affected and unaffected at the *CTG18.1* locus, there was no evidence for linkage between *CTG18.1* and BPAD. In six BPAD families moderately enlarged $(CTG)_{>86}$ alleles were evident. PCR detected alleles ranging from $(CTG)_{10}$ to $(CTG)_{250}$. Alleles between 10 and 37 repeats were inherited stably in 261 out of 262 transmissions, whereas moderately enlarged (53-250 repeats) alleles demonstrated instability in 18 transmissions observed. Enlarged alleles at the *CTG18.1* locus segregated randomly with respect to BPAD. Sequencing across the enlarged repeats confirmed that the expansion derived from the CTG repeat and that the CTC repeat remained stable. Further research indicated very enlarged alleles (800-2100 repeats) at the *CTG18.1* locus that could only be detected by genomic Southern blots in three BPAD individuals and a Danish pedigree. However, these very enlarged alleles also segregated independently of the BPAD phenotype.

In the same study it was reported that some CEPH families had expanded Repeat Expansion Detection (RED) products mapped to 18q (Schalling *et al.*, 1993). CEPH families 1334, 1344 and 1420 were amplified by PCR at locus *CTG18.1*. Families 1344 and 1420 inherited enlarged alleles in a mendelian manner and 1334 had normal stable alleles. This illustrated that the RED technique (Figure 1.2) may have detected the expanded alleles at the *CTG18.1* locus in families 1344 and 1420, but that another locus may be responsible for the RED bands detected in family 1334. Typing at the *CTG18.1* locus in individuals from BPAD families found an expansion frequency (>40 repeats) of 3% (Verheyen *et al.*, 1999),

confirming earlier findings (Breschel *et al.*, 1997). Alleles greater than 100 repeats were not accurately sized and so could not be commented on in terms of instability in transmission. However, a contraction of an approximately 133 repeat allele to 87 repeats in maternal transmission was observed.

The *CTG18.1* locus has also been investigated in cases of familial Parkinson's disease (PD) in combination with the RED method (Schraen-Maschke *et al.*, 1999). Expanded CAG/CTG RED products in some PD patients (21.9%) led to screening at the *CTG18.1* locus and two expanded alleles of 65 and 67 repeats were detected in a patient and control, respectively. *CTG18.1* was discounted of involvement in the PD phenotype.

1.4 Locus *ERDA1*

A second expandable CAG/CTG repeat not associated with a disease was found on chromosome 17q21.3 (Nakamoto *et al.*, 1997). As with *CTG18.1*, this repeat had been detected by the RED method and subsequently cloned. The DNA fragment was named *ERDA1* (expanded repeat domain CAG/CTG 1) and repeats ranged from 10-30 and 55-92 in healthy Japanese and Caucasian individuals, demonstrating a bimodal distribution. The frequencies of the larger *ERDA1* alleles are again higher than in DM. 26% of Japanese and 8% of Caucasian alleles had greater than 51 repeats, with alleles from 66-87 repeats stably transmitted from parent to offspring. The larger alleles possess an A→C polymorphism immediately 3' to the CTG repeat. An independent study also directly compared allele distributions in Japanese and French individuals, dividing the bimodal distribution into short and long alleles and naming the locus *Dir1*. Japanese alleles ranged from 10-26 and 52-92 repeats and French alleles ranged from 10-25 and 50-78 repeats with heterozygosities of 85% and 82% respectively (Ikeuchi *et al.*, 1998). The study agreed with the previous study that large alleles are more prevalent in the Japanese (38.2%) than in French (17.2%). Maternal transmissions of large alleles were more unstable with a bias toward expansions (7/15 expansions, 3/15 contractions) and paternal transmissions were biased toward contractions (4/11 expansions, 5/11 contractions) disagreeing with the earlier study. However the Nakomoto study looked at very few transmissions. Northern and RT-PCR analysis suggests there may be expression of the *Dir1* fragment in tissues including heart, skeletal muscle and pancreas, though the CAG/CTG repeat sequence was included in the probe and hence may detect other expressed transcripts including repeats. Both studies noted that *ERDA1* could be

responsible for some of the expanded bands detected by the RED method and should be taken into consideration.

The bimodal distribution of *ERDA1* alleles was confirmed in several different isolated populations including Nigerians, Papua New Guinea Highlanders, Chinese and CEPH reference families (Deka *et al.*, 1999). Nigerians had a greatly reduced frequency of alleles greater than 40 repeats (5.9%) compared to other populations (average 15%). Further investigation uncovered interruptions in the Nigerian small alleles of CAT and CAC repeats and the A polymorphism 3' of the repeat, which presents as a C in expanded alleles, implying that these small alleles would possess enhanced stability and hence rarely give rise to expanded alleles in the population. When observing transmissions at this locus, there was no parental effect on the transmission of expanded alleles, however there was a difference in the nature of paternal mutations (45.9% mutation rate of 61 transmissions of large alleles, 93% of which are contractions), compared to maternal mutations (47.1% mutation rate of 34 transmissions of large alleles, 75% of which are expansions). Verheyen *et al.* (1999) also found an expansion (>40 repeats) frequency of 19% in BPAD families, with instability of 65 and 84 repeat alleles in transmission. This differed from Nakamoto *et al.* (1997) and aspects of the study by Deka *et al.* (1999) but agreed with Ikeuchi *et al.* (1998) in that the maternal transmission of the 65 repeat allele gave rise to expansions (67-71 repeats) and the paternal transmission of the 84 repeat allele gave rise to contractions (73-74 repeats). Expanded RED products in PD patients prompted a screen at the *ERDA1* and *CTG18.1* loci which concluded that 100% of expanded RED products could be attributal to these two loci (Schraen-Maschke *et al.*, 1999). The largest *ERDA1* allele detected in PD families was 75 repeats, although normal alleles were also detected in patients and the 75 repeat allele was present in the unaffected daughter of an affected patient (Schraen-Maschke *et al.*, 1999). Hence *ERDA1* and *CTG18.1* were discounted as contributing to the PD phenotype.

1.5 Triplet Repeats and Psychiatric Disorders

Affective Disorder is present as a spectrum of phenotypes including Bipolar Affective disorder (BPAD) affecting 1% of the population and Unipolar affective disorder (UPAD) affecting 6% of the population. BPAD can be subdivided into types I and II, consisting of mood swings from depression, such as low mood, sleep and appetite disturbance, psychomotor abnormalities, fatigue, diminished self-attitude and cognitive slowing, and in type I, mania such as elation, impulsivity, increased energy and activity and rapidity of speech and cognition. Psychiatric disorders such as BPAD are thought to be multigenic due to the

variability in symptoms, even within families and the mode of inheritance remains unclear, although mothers are more likely to transmit the phenotype with a relative lack of male to male transmission (Gelernter, 1995). Family studies of affected probands conclude an increased risk of UPAD, major depressive disorder (MDD), cyclothymia (CY) and schizoaffective (SA) disorders to a first degree relative (Verheyen *et al.*, 1999). BPAD has been linked to chromosomes 1q31-q32 (Detera-Wadleigh *et al.*, 1999), 4p16 (Blackwood *et al.*, 1996), 7q31 (Detera-Wadleigh *et al.*, 1999), 11 (McInnes *et al.*, 1996), 12q23-24 (Ewald *et al.*, 1998), 13q32 (Detera-Wadleigh *et al.*, 1999), 18 (Berrettini *et al.*, 1994, Stine *et al.*, 1995, Coon *et al.*, 1996, De bruyn *et al.*, 1996, Freimer *et al.*, 1996, McInnes *et al.*, 1996, McMahon *et al.*, 1997, Detera-Wadleigh *et al.*, 1999), 21 (Straub *et al.*, 1994 Detera-Wadleigh *et al.*, 1996), 22q11-q13 (Detera-Wadleigh *et al.*, 1999) and X (Baron *et al.*, 1987). In particular, chromosome 18q has been the focus of detailed linkage analysis. Stine *et al.* (1995) demonstrated linkage at 18q21 in combined BPAD type I and II individuals, McInnes *et al* (1996) found highest linkage in two combined families with BPAD type I at 18q22-q23 and McMahon *et al.* (1997) found affected sib pairs had excessive allele sharing at 18q markers, D18S38 demonstrating the strongest linkage, but other studies have failed to detect linkage of BPAD to chromosome 18 (Maier *et al.*, 1995, Pauls *et al.*, 1995, Detera-Wadleigh *et a.,* 1997)

Schizophrenia is characterised by gross disturbances of perception (hallucinations) and cognition (delusions and thought disorder) as well as the negative symptoms of amotivation and paucity of thought (McInnis and Margolis, 1998), affecting approximately 1% of the population. The disease has been linked to chromosomes 2q (Levinson *et al.*, 1998), 3p (Pulver *et al.*, 1995), 4q (Levinson *et al.*, 1998), 5q (Straub *et al.*, 1997), 6p (Moises *et al.*, 1995, Wang *et al.*, 1995), 8p (Pulver *et al.*, 1995), 9 (Moises *et al.*, 1995, Levinson *et al.,* 1998), 10q (Levinson *et al.*, 1998), 11q (Levinson *et al.*, 1998), 18q (Hovatta *et al.*, 1999), 20 (Moises *et al.*, 1995) and 22q (Coon *et al.*, 1994, Pulver *et al.*, 1994), in genomewide scans of affected families. The onset of the disease and the outcome is worse in males and anticipation has been observed in several different studies (Asherson *et al.*, 1994, Chotai *et al.*, 1995, Thibaut *et al.*, 1995, Bassett and Husted, 1997). As described earlier, anticipation is a decreasing age of onset with increasing severity of symptoms within a pedigree of a disease family. These findings implicate expandable triplet repeats as the genetic mutation giving rise to schizophrenia. Hence, the RED (Repeat Expansion Detection) method (Schalling *et al.*, 1993) (Figure 1.2) has been utilised to look for evidence of triplet repeat expansion (CAG/CTG in most cases due to the higher frequency of these expanded alleles contributing to disease phenotypes) in various psychiatric disorders such as BPAD and schizophrenia, with conflicting data.

Either a higher percentage or larger expanded RED products have been found in BPAD (Lindblad *et al.*, 1995, Mendlewicz *et al.*, 1997, Verheyen *et al.*, 1999) and schizophrenia (Morris *et al.*, 1995, Burgess *et al.*, 1998, Vincent *et al.*, 1998). In cases of BPAD, ethnicity (Lindblad *et al.*, 1995), sex of transmitting parent (Mendlewicz *et al.*, 1997), type I or II symptoms (Lindblad *et al.*, 1995) and age of onset (Mendlewicz *et al.*, 1997, Verheyen *et al.*, 1999) have been linked with RED product size. Belgian type I affecteds have significantly larger RED products than Swedish or type II patients (Lindblad *et al.*, 1995). Increasing severity of symptoms (Mendlewicz *et al.*, 1997, Verheyen *et al.*, 1999) and a link between maternal but not paternal transmission and expansion of RED product (Mendlewicz *et al.*, 1997) has been observed. Other studies could not replicate the positive RED findings in BPAD patients (Vincent *et al.*, 1996, Li *et al.*, 1998, Verheyen *et al.*, 1999, Vincent *et al.*, 1999). Larger RED products have also been found in schizophrenia patients (Morris *et al.*, 1995, Burgess *et al.*, 1998, Vincent *et al.*, 1998) with sex effects in females (Morris *et al.*, 1995) or males (Burgess *et al.*, 1998). Twins fully or partially discordant for the disease also demonstrated larger RED products than those discordant (Vincent *et al.*, 1998), but no difference in RED product size between affected and unaffected has been found in other studies (Vincent *et al.*, 1996, Laurent *et al.*, 1998, Li *et al.*, 1998).

Following on from this, groups have screened many repeat loci in order to find an expandable locus for BPAD and schizophrenia. One strong candidate for schizophrenia seemed to be the *hSKCa3* gene. *hSKCa3* contains 2 CAG repeats, one of which is highly polymorphic (Chandy *et al.*, 1998). Initial reports located *hSKCa3* to chromosome 22q11, near a region linked to schizophrenia. Allele typing found that alleles >19 repeats were more commonly found in the schizophrenia group compared to controls (Chandy *et al.*, 1998), strengthened with results from Bowen *et al.* (1998) and Dror *et al.* (1999). Then *hSKCa3* was shown to map to 1q21 and not 22q11 (Austin *et al.*, 1999, Dror *et al.*, 1999) and the point was made that the alleles at the *hSKCa3* locus were not large enough to account for RED expansions observed in patients. Further work demonstrated that there was either excessive transmission of the smaller alleles within affected families or lack of transmission of the larger alleles (Li *et al.*, 1998, Stober *et al.*, 1998, Austin *et al.*, 1999, Wittekindt *et al.*, 1999) or no significant difference between allele sizes in BPAD, shizophrenia or control individuals (Hawi *et al.*, 1999). A *Bal*I restriction polymorphism (changing Serine to Glycine) in exon 1 of the *dopamine D3 receptor* gene in conjunction with a telomeric dinucleotide repeat marker (D3S1310) has also been investigated in schizophrenia patients (Jonsson *et al.*, 1999). Many previous reports had associations between *DRD3* genotypes with various subtypes including males, individuals who respond to medication, individuals who have a family history of the disease, suffer alcohol or substance abuse or have been diagnosed with disorganised or undifferentiated shizophrenia. The subsequent study found no difference between the *Bal*I

genotypes of schizophrenics or controls (Jonsson *et al.*, 1999). At D3S1310 however, shizophrenics had a lower frequency of the 141bp allele and a higher frequency of the 143bp allele compared to controls. Homozygosity for the 143bp allele was the most frequent genotype in both patients and controls, but at a higher frequency in schizophrenics (Jonsson *et al.*, 1999). No evidence for linkage disequilibrium between this polymorphism and marker D3S1310 was found but now another study has uncovered several polymorphisms in novel exons 5' to the named exon 1 which are currently being tested (O'Donovan and Owen, 1999).

Distribution of the polymorphic B33 CTG repeat (9-18 repeats) on chromosome 3 showed no difference between periodic catatonia (subtype of schizophrenia) patients and controls (Bengel *et al.*, 1998) although an increased percentage heterozygosity was observed. Schizophrenic and BPAD patients were typed at a polymorphic CAG repeat (6-9 repeats) in the *IL9 receptor* gene at Xq28/Yq21, showing a slight increase in the frequency of the larger of the two most common alleles (8 repeats) in patients, but not enough to be significant (Hawi *et al.*, 1999). Separating the samples based on sex increased the significance of association of the 8 repeat allele with BPAD in males, but not in females. No association was found between any allele and schizophrenia (Hawi *et al.*, 1999). The polymorphic TTA repeat (122-143bp) in the *phospholipase A2 group 1B (PLA2G1B)* gene has also been investigated in schizophrenics (Strauss *et al.*, 1999). No difference was found between the distribution of the alleles in Caucasian or African-American controls compared to patients, the most frequent alleles being 122/125bp (Caucasian) and 125/134/137bp (African-American). The imperfect polymorphic CAG repeat in the *TATA-binding protein (TBP)* gene at 6q27 (Imbert *et al.*, 1994) was typed in schizophrenia or BPAD patients and controls (Rubinsztein *et al.*, 1996). No difference was found between different ethnic groups (Caucasians, South African blacks, Sub-Saharan Africans, Asian Indians) and schizophrenia or BPAD patients. Furthermore, over 50 CAG/CTG loci distributed on autosomal (44) and X (7) chromosomes have been screened by allele typing with fluorescence and radioisotope methods after the indication that there could be a sex effect in BPAD (O'Donovan *et al.*, 1995). All loci were excluded for association with BPAD based purely on size as the largest allele detected was only 37 repeats (Guy *et al.*, 1997).

During the course of investigations into the possibility of triplet repeat expansion causing BPAD or schizophrenia using the RED technique, both the *CTG18.1* and *ERDA1* loci were published. This presented novel expandable trinucleotide repeat loci as possible candidates for BPAD and /or schizophrenia. Psychiatric patients were screened using the RED method with a CTG probe and though large RED products were evident, PCR allele typing at *CTG18.1* and *ERDA1* pointed to 94% of trinucleotide repeat expansions detected by RED being due to expansions at these loci (Sidransky *et al.*, 1998). Again, screening large RED

products in BPAD individuals found that 75% were due to *ERDA1* compared to 83% in controls (Lindblad *et al.*, 1998). In total, 8 out of 9 individuals with large RED products had either *CTG18.1* or *ERDA1* expansions. At the *CTG18.1* locus, 19% of affected individuals had an expansion, compared to only 8% of controls with a significant difference between affected and unaffected in RED products, producing a negative correlation between *CTG18.1* repeat length and age of onset in families. Therefore, it is important to note that if both *CTG18.1* and *ERDA1* can be excluded as the cause of a large RED product, the remaining large products are likely to indicate the disease allele. In a study on cases of schizophrenia which have demonstrated expanded RED products, *CTG18.1* was screened but not all RED products could be attributed to this locus (Burgess *et al.*, 1998) and *ERDA1* has yet to be investigated. Twins either cordant or discordant for schizophrenia were also analysed at *CTG18.1* but showed no differences in repeat size, demonstrating that this locus alone is not responsible for the variation in RED product sizes between siblings (Vincent *et al.*, 1998).

Verheyen *et al.* (1999) reanalysed Belgian BPAD blood samples published by Lindblad *et al.* (1995) which had demonstrated that Belgian type I affecteds had significantly larger CAG/CTG products than controls. Both BPAD patient and control samples that had expanded RED products were typed at both *CTG18.1* and *ERDA1*, with 91.3% and 81.5% respectively having alleles >40 repeats at either of these loci. For *ERDA1*, the overall allele distributions were significantly different between BPAD affecteds and controls. Although the distribution of the largest *ERDA1* alleles in individuals was not different, the proportion of BPAD individuals with one or two expanded alleles was significantly larger in the BPAD affecteds and the number of large (>40 repeats) alleles was higher in BPAD affecteds than controls. For *CTG18.1*, there was no significant difference between the allele sizes in BPAD individuals or controls. *CTG18.1* and *ERDA1* loci were typed in BPAD families, in both affected and unaffected individuals (Verheyen *et al.*, 1999). 90.2% of expanded RED products were accounted for by expansions at *CTG18.1* or *ERDA1*. *CTG18.1* was discounted as a BPAD gene in a particular family which had previously shown linkage to 18q21.33-q23 (De bruyn *et al.*, 1996) as they did not demonstrate expansions at this locus (Verheyen *et al.*, 1999). However, a slight negative correlation between *ERDA1* allele size and age of onset was observed (Verheyen *et al.*, 1999). A study of *CTG18.1* and *ERDA1* allele sizes in caucasian BPAD type I patients and controls however, showed no significant difference between the two groups, either in distributions of maximum allele sizes, the frequency of large alleles (>40 repeats) or between age of onset and frequency of large alleles (Guy *et al.*, 1999). Subsequent comparison of RED product size and alleles at the *CTG18.1* and *ERDA1* loci revealed only approximately 41% can be attributed to these loci, a much lower estimation than previous studies (Lindblad *et al.*, 1998, Sidransky *et al.*, 1998, Verheyen *et al.*, 1999). Analysis of individuals with familial depression revealed an expanded *ERDA1*

allele present in three generations of one family (Vincent *et al.*, 1999). Combined symptoms of the maternal grandparents included schizophrenia, drug abuse and major depression. The proband had inherited an expansion from 102 to 115 repeats at the *ERDA1* locus and also suffered major depression, anxiety, abuse and dysthymic disorder. Her children inherited 82 and 115 repeats respectively but there was no data available on their health (Vincent *et al.*, 1999).

Given the growing number of diseases associated with trinucleotide repeats and the reports of anticipation in psychiatric disorders it seems likely that dynamic mutations could be a contributory factor. Although BPAD and schizophrenia are thought to be multigenic disorders, screening of novel trinucleotide repeat loci has been undertaken to try to find a causative mutation. The RED method gives an indication of the presence of an expandable trinucleotide repeat in the genome, but not its location. The RED method does seem to point to an increased frequency of expanded repeats in psychiatric patients and *CTG18.1* and *ERDA1* are ideal candidates. However, although there are also reports of increased expanded alleles at these loci in affected individuals, there is no absolute correlation between expansions and disease, unlike *DM-1* and *HD*.

1.6 The *SEF2-1* gene and its protein function

Discovery of the *CTG18.1* locus characterised an expandable CTG repeat putatively located in the third intron of the human *SEF2-1* gene at chromosome 18q21.1 (Breschel *et al.*, 1997). SEF2-1 is one of a family of nuclear proteins designated SL3-3 enhancer factors 2 belonging to the basic helix-loop-helix (B-HLH) class of DNA-binding proteins. These proteins interact with an Ephrussi box (E box)-like motif (-NNCANNTGNN-) which gives them the name E proteins. E proteins can be divided into 3 types, ubiquitously expressed, tissue-specific and inhibitors of DNA-binding. Class I E proteins are ubiquitously expressed and bind as homodimers or heterodimers with a Class II protein to an E2 box (-G/ACAGNTGT/G-) which have been identified in muscle and pancreas-specific enhancers. All Class I proteins are structurally related and well conserved in the HLH domain and include the proteins E12 and E47 derived from the E2A gene, HEB, SEF2-1 and daughterless of *Drosophila* which is involved in the control of sex determination and neurogenesis. Class II E proteins are tissue-specific and do not bind as homodimers to an E2 box. Members of Class II E proteins include MyoD, myogenin, myf-5, herculin, proteins that induce myogenesis and *Drosophila* gene products involved in neurogenesis including members of the *achaetescute* gene family (Bain *et al.*, 1993). Id proteins are inhibitors of DNA binding

and bind to either Class I or II E proteins as heterodimers. Id proteins lack the basic DNA binding domain and are known to bind with the least affinity to any of the E proteins.

SEF2-1 is also known as *ITF-2* (Immunoglobulin Transcription Factor 2), (Henthorn *et al.*, 1990) in human, *E2-2* in mouse (Corneliussen *et al.*, 1991) and *TFE* in dog (Javaux *et al.*, 1991). *ITF-2* was discovered in 1990 (Henthorn *et al.*, 1990) while investigating the Immunoglobulin heavy and κ light chain gene enhancers and how they are activated by ubiquitous and developmentally regulated proteins. Two cDNAs were isolated which encoded the proteins named ITF-1 (E2A) and ITF-2 which were expressed in a variety of cell types and bound to the μE5/κE2 motif found in both enhancers. ITF-1 and ITF-2 are structurally and functionally similar, but derive from separate genes. ITF-1 and ITF-2 interact with one another as proteins through their helix-loop-helix motifs and each possess a domain in the amino terminus that dictates transcription activation.

The *SEF2-1* gene encodes four main alternative transcripts, SEF2-1A-D (although 10 different cDNAs may have been identified (Hermann *et al.*, 1998)), identical except in their 5' ends. SEF2-1A and B are the most prevalent transcripts, SEF2-1B being full length, SEF2-1A and SEF2-1D being shorter and containing an additional amino acid motif known as the RSRS domain 5' to the bHLH domain (Corneliussen *et al.*, 1991). MITF-2A and B are the mouse homologues of SEF2-1A and B, but MITF-2B contains the RSRS domain (Skerjanc *et al.*, 1996). MITF-2B seems to downregulate MyoD induced expression of the cardiac α-actin protmoter whereas MITF-2A has no effect, but removal of the RSRS domain suggests this motif is not responsible for the inhibition activity (Skerjanc *et al.*, 1996) and in fact the alternative 5' ends are responsible. It is not yet clear what role the RSRS domain may play in DNA-binding or protein-protein interactions, especially as the closest relatives of MITF-2, HEB and E2A do not contain a RSRS domain. MITF-2B transcripts have been detected in all cell types including stem cells, cardiac muscle, skeletal muscle, neuroectoderm, mouse brain, mouse liver, mouse leg muscle and heart, all in equal amounts. MITF-2A is present in all samples that contain neurons such as neuroectoderm, skeletal muscle and brain, but at variable levels in other tissues.

SEF2-1 has also been implicated in the expression of Somatostatin Receptor 2 (SSTR2), one of a family of 7-helix transmembrane spanning receptors which regulate potassium and calcium channels and cytoplasmic signal transduction pathways (Rens-Domiano and Reisine, 1992). Somatostatin receptors are highly expressed in the hypothalamic-hypophyseal system, regulating the adenohypophyseal release of growth hormone (GH), thyroid stimulating hormone (TSH) and prolactin (Brazeau *et al.*, 1972), the somatostatinergic system in the

central nervous system, modulating cognitive and vegetative functions (Epelbaum *et al.*, 1994) and the gastroenteropancreatic system of endocrine cells regulating gastric acid and gastroenteropancreatic hormone release (Nelson-Piercy *et al.*, 1994). SSTR2 and SSTR4 are also expressed in the murine embryo brain (Maubert *et al.*, 1994). By screening a murine brain expression library with a probe made up of repeated units of the SSTR2 promoter, MITF-2B was found to bind to an E box within 90bp of the ATG start site (Pscherer *et al.*, 1996). Direct comparison of the submitted cDNA sequences from Skerjanc and Pscherer has revealed some discrepencies in MITF-2B. The definitions of the different isoforms of ITF-2/SEF2-1/E2-2 includes whether the RSRS domain is present, identical 3' ends and differing 5' ends. However, MITF-2B as sequenced by Pscherer does not contain the RSRS domain and has differing 5' and 3' ends. MITF-2B defined by Skerjanc seems to be the correct mouse homologue of SEF2-1B as published by Corneliussen. Further expression work on SEF2-1 was carried out with a human SEF2-1 probe, detecting a 7.5 Kb transcript in brain, heart, placenta, skeletal muscle and lung of the human adult, heart, lung and highly in brain of the human embryo, heart, brain, lung and skeletal muscle in mouse and in whole murine embryo (Pscherer *et al.*, 1996). Gel retardation assays demonstrated that SEF2-1 GST fusion protein alone binds to the E box, but requires at least a second protein endogenous to the cell type to activate transcription of SSTR2. TFIIB did form a second complex with SEF2-1 when it was bound to DNA, raising the possibility that SEF2-1 can initiate transcription by recruiting transcription factors to the promoter of SSTR2 (Pscherer *et al.*, 1996). Further characterisation of proteins binding to the SSTR2 promoter uncovered MIBP1, a sequence-specific DNA-binding zinc finger which binds to a TC-rich sequence directly upstream of the E box and enhances expression (Dorflinger *et al.*, 1999). MIBP1 was found to directly interact with the SEF2-1 protein and the specific bHLH domain was responsible. MIBP1 did not interact with either E12 or E47 (alternatively spliced products of the E2A gene), suggesting a possible role for the RSRS domain in protein interactions. Furthermore, MIBP1 demonstrated an identical expression pattern with SEF2-1 and SSTR2 in murine heart, skeletal muscle and brain, specifically the cerebra cortex, hippocampus and corpora amygdala (Dorflinger *et al.*, 1999).

SEF2-1 protein has been hypothesised to work in conjunction with HEB and E2A proteins in the development of B-cells (Zhuang *et al.*, 1996). Mice homozygous for SEF2-1 or HEB mutations can produce B-cells, but have reduced numbers (50%) of pro-B cells indicating both genes are not essential for the establishment of B-lymphocytes. Homozygous lack of SEF2-1 and heterozygous lack of E2A results in embryonic lethality. It was suggested that SEF2-1 and HEB may form heterodimers with Id E proteins which inhibit DNA binding. By preventing the Id proteins binding to a third E protein, E2A, it is left to form homodimers and

bind to an enhancer or E box in the promoter region of a gene(s) controlling B lymphocyte development (Zhuang *et al.*, 1996).

As SEF2-1 is ubiquitously expressed it is likely that it plays many roles in the differentiation and development of various tissues. Increased expression of SEF2-1 has been discovered in prostate cancer cells of three different patients compared to normal tissue (Liu *et al.*, 1997). SEF2-1 lacking RSRS binds to a sequence in the brain-specific FGF-1.B promoter and represses expression (Liu *et al.*, 1998) whereas expression is not affected if the RSRS is present. An interesting observation was that the human brain contains proportionally more SEF2-1 with the RSRS domain than without (Liu *et al.*, 1998) which is likely to be due to observed elevated expression of SEF2-1D (Skerjanc *et al.*, 1996). Calcium loaded calmodulin has also been implicated in regulating SEF2-1 binding to DNA (Hermann *et al.*, 1998), and interaction is determined by sequence directly 5' to the bHLH domain. Could this be a function of the RSRS domain?

1.7 Genomic instability in Cancer

1.7.1 Hereditary Non-polyposis Colon Cancer (HNPCC)

HNPCC can affect as many as 1 in 200 individuals and accounts for 3-6% of all colon cancers. An HNPCC kindred is defined by the Amsterdam criteria as those in which three or more closely related family members in at least two successive generations have had histologically verified diagnosis of colorectal cancer, at least one of whom was diagnosed before 50 years of age (Vasen *et al.*, 1991). The Amsterdam criteria has now been revised to incorporate HNPCC-associated cancers of the endometrium, small bowel, ureter or renal pelvis in addition to the colon (Vasen *et al.*, 1999). HNPCC had been linked to chromosome 2p16-15 (Peltomaki *et al.*, 1993) and 3p21 (Lindblom *et al.*, 1993) and is often divided into two subgroups, type I and type II. Type I have a hereditary predisposition to colorectal cancer that is distinguished from sporadic colorectal cancer by a younger age of onset and the carcinomas are often observed in the proximal colon. Type II patients have an increased risk for cancers of the uterus, ovary, breast, stomach, kidney, small intestine, skin and larynx in addition to the colon. Significant changes in (CA) dinucleotide allele lengths were observed more frequently in HNPCC than sporadic tumours at six different loci, suggesting the underlying cause of HNPCC could be a defect in mismatch repair (Aaltonen *et al.*, 1993).

The majority of HNPCC cases are now known to be caused by inherited germline mutations in one of five DNA mismatch repair enzymes *hMLH1* (Bronner *et al.*, 1994, Papadopoulos *et al.*, 1994, Wheeler *et al.*, 1999), *hMSH2* (Leach *et al.*, 1993), *hPMS1* (Nicolaides *et al.*, 1994), *hPMS2* (Nicolaides *et al.*, 1994) or *hMSH6* (Akiyama *et al.*, 1997, Miyaki *et al.*, 1997). The role of the DNA mismatch repair (MMR) pathway in conjunction with the nucleotide excision repair (NER) pathway is to detect mismatches that occur during replication and correct them by excising the incorrect strand and replacing it with a new, correctly synthesised strand. Extensive research has been carried out to characterise the MMR proteins in *E.coli* and yeast to precisely define their function, however even though the respective homologues have been found in humans, their precise roles are elusive. For example, in *E.coli* the mismatch repair pathway is well defined and involves the co-operation of several proteins known as the "Repairosome" (Figure 1.3). Within the repairosome, it is known that MutS recognises and binds to the mismatch in the DNA (Su and Modrich, 1986), with MutH binding to hemimethylated GATC sites on the incorrect strand (Welsh *et al.*, 1987). MutL is then recruited by MutS, inactivating MutH (Grilley *et al.*, 1989). MutU (UvrD helicase) acts in conjunction with exonucleases (RecI and ExoI) to excise the unmethylated strand and the new strand is now synthesised in place of the old, incorrect strand by DNA polymerase III, single-strand DNA-binding protein and DNA ligase (Matson and George, 1987, Lahue *et al.*, 1989, Modrich, 1989, Modrich, 1991, Cooper *et al.*, 1993). In yeast, there are seven homologues of *MutS* (*MSH1-6* and *HSM3*) and four *MutL* homologues (*MLH1-3* and *PMS1*) (Kirkpatrick, 1999). MSH1-3 and 6 are involved in mismatch repair and MSH4 and 5 are involved in chromosomal crossover and segregation (Kirkpatrick, 1999). No *MutH* homologues have been defined but yeast does not methylate its DNA and hence this protein would not have a function.

The human homologue of *MSH2* (*hMSH2*) is made up of 16 exons, cloned using degenerate primers to conserved sequences of the known *MutS* homologues and localised to chromosome 2 (Fishel *et al.*, 1993). As some HNPCC families had been linked to chromosome 2p16-15 (Peltomaki *et al.*, 1993) *hMSH2* was investigated as a potential candidate gene. Only 48% of the *MSH2* sequence was examined for mutations, but 29% of the 27% of tumours that demonstrated dinucleotide repeat instability had a transition of T to C in this gene. Analysis of only exon 13 of *hMSH2* revealed two mutations of T to C transition and a G deletion in four separate families (Jeon *et al.*, 1996). Liu *et al.*(1996) noted large *hMSH2* genomic deletions in other patients encompassing exons 4-8 and 1-6. HNPCC families that did not have nucleotide changes in *hMSH2* were investigated further and four different deletions covering parts of exons 1, 2, 3 and 6 were uncovered, but only in Dutch and not Norwegian kindreds (Wijnen *et al.*, 1998). In total, 24% of HNPCC families defined by

the Amsterdam criteria have a *hMSH2* mutation, 27% of which are deletions and 73% are point mutations (Wijnen *et al.*, 1998).

hMLH1, hPMS1 and *hPMS2* were discovered by scanning a database for sequences with the highest similarity to yeast and bacterial mutator genes, designing PCR primers and localising the genes to chromosomes 3, 2 and 7 respectively (Papadopoulos *et al.*, 1994). *hMLH1* was further localised to 3p21.3 and shown to be expressed in colon, lymphocytes, breast, lung, spleen, testes, prostate, thyroid, gall bladder and heart (Papadopoulos *et al.*, 1994). Due to the previous linkage of HNPCC families to 3p21 (Lindblom *et al.*, 1993), *hMLH1* cDNA from normal and HNPCC individuals was sequenced and a deletion of a full exon was found in affected Finnish individuals with a common ancestor. Other pathogenic mutations included 4bp deletions and insertions giving rise to frameshifts, altering the carboxy terminus which is identical in yeast and humans, suggesting it has a very important role. In all, 90% of families linked to 3p had *hMLH1* mutations (Nicolaides *et al.*, 1994). Bronner *et al.* (1994) also isolated *hMLH1* by using degenerate primers to conserved sequences in bacteria and yeast *MutL* genes. They noted that the last 13 amino acids are identical between yeast and human *MLH1*. *hMLH1* was mapped to 3p21.3-23, and a C to T transition detected in an affected family, hypothesised to interfere with the conformation of the protein. Han *et al.* (1995) published that the *hMLH1* gene sequence is made up of 19 exons over 100kb and detected misense, frameshift and splice altering mutations in type II patients who also developed cancers of the pancreas, endometrium, ovary or skin. Liu *et al.* (1996) detected a C to T transition in *hMLH1* in one patient heterozygous in normal tissue and homozygous in cancerous colon. Other mutations included deletion of exon 15 and G to A transitions in intron 14 and exon 15. Again, the alleles were heterozygous in normal tissue but homozygous in cancerous tissue. This implies that the mechanism of tumour development involves inheritance of a mutated allele and loss of function of the second allele occurring through somatic mutation. Ten *hMLH1* and 4 *hMSH2* mutations were detected in 15 families from another study, ranging from transitions to deletions of whole exons (Liu *et al.*, 1996). Patients screened for *hMLH1* and *hMSH2* mutations demonstrate that 38% of *hMLH1* mutations are in exons 15 and 16 (Wijnen *et al.*, 1997). 49% of families fulfilling the Amsterdam criteria have mutations in *hMLH1* or *hMSH2* compared to only 8% of families where one or more of the criteria were not fulfilled.

Inter-ethnic differences can alter the frequency of *hMLH1* and *hMSH2* mutations in HNPCC families. 24% of HNPCC in Japanese is due to *hMLH1* mutations (Han *et al.*, 1995), 31% and 33% of cases in North America are due to *hMSH2* and *hMLH1* respectively (Liu *et al.*, 1996), whereas in Finland 83% of affected families are due to *hMLH1* and only 3% due to *hMSH2*, probably attributed to a founder effect (Nystrom-Lahti *et al.*, 1996). Chan *et al.*,

(1999) detected an amino acid change in exon 9 of *hMSH6* and nucleotide changes and insertions mainly resulting in stop codons have been detected in exons 2, 4, 8 and intron 10 (Wu *et al.*, 1999). Haplotype analysis of flanking single nucleotide polymorphisms in patients with a *hMSH6* T insertion in exon 4 also suggests a founder effect (Wu *et al.*, 1999). *hPMS1* and *hPMS2* have been subsequently localised to 2q31-33 and 7p22 respectively (Nicolaides *et al.*, 1994) and in 40 HNPCC individuals screened for MMR mutations, 45% were in *hMLH1* or *hMSH2* and in the remaining 55%, a C to T transition was discovered in *hPMS1* and a deletion in *hPMS2* (Nicolaides *et al.*, 1994) but not all mutations were accounted for. Nucleotide substitutions and deletions in MMR genes do not account for all HNPCC cases and in individuals with as of yet unidentified sequence alterations, hypermethylation of the *hMLH1* promoter has been discovered in tumour samples, resulting in no expression of the protein (Wheeler *et al.*, 1999). To date, hypermethylation of *hMSH2, hPMS1, hPMS2* or *hMSH6* has not been detected.

Once heterozygosity of a MMR gene is lost, loss of post-replicative DNA mismatch repair increases mutation rates 100-600 times above normal levels (Bhattacharyya *et al.*, 1994, Eshleman *et al.*, 1995). This mutator phenotype is described as replication error positive (RER+) with the appearance of microsatellite instability (MIN) and a rapid accumulation of mutations in genes important for cell growth which leads to tumour development. RER- cancers tend to be aneuploid, demonstrating large alterations of the genome and mutations in *APC, KRAS* and *p53* are common. RER+ tumours are diploid, do not show major chromosomal rearrangements and very few mutations in *APC, KRAS* or *p53*. These two different types of genomic instability suggest there are two distinct pathways of progression from premalignant to carcinoma cells (Eshleman and Markowitz, 1996).

It has been postulated that 75-100% of HNPCC tumours demonstrate microsatellite instability whereby the size of alleles differs between normal and cancerous tissue from the same individual (Aaltonen *et al.*, 1993, Lindblom *et al.*, 1993, Aaltonen *et al.*, 1994), In fact, direct comparison of sporadic and HNPCC tumours found 83% of sporadic were MIN- (stable) and 85% of HNPCC were MIN+ (Pedroni *et al.*, 1999). Ionov *et al.* (1993) and Thibodeau *et al.* (1993) demonstrated that HNPCC tumours and some sporadic colorectal cancers exhibit widespread microsatellite instability on different chromosomes. Dinucleotide repeats (usually CA) are the most frequent type of repeat to be analysed due to their availability in the genome (estimated to be about 50,000-100,000) and their relative stability, particularly in transmission.

Two different dinucleotide (CA) and mononucleotide (A) repeats were analysed in HNPCC individuals with 92% displaying microsatellite instability and 81% of those were unstable at all four loci studied (Liu *et al.*, 1996). Wijnen *et al.* (1997) analysed seven dinucleotide, two

trinucleotide (including the *DRPLA* locus) and two tetranucleotide repeats in blood and tumour DNA from 24 patients. 27% of AMS- (families not fulfilling Amsterdam criteria) and 67% of AMS+ (families fulfilling Amsterdam criteria) showed microsatellite instability, but of the AMS-, MMR mutations could not be detected. Patient groups split into HNPCC, symptoms similar to HNPCC and sporadic demonstrated that sporadic cancers have decreased levels of microsatellite instability at less loci and patients with *hMLH1* mutations had a higher rate of MIN than patients with *hMSH2* mutations (Lamberti *et al.*, 1999). Observations from patients with *hMSH6* mutations showing low or no microsatellite instability led Wu *et al.* (1999) to evaluate 15 markers in tumours of this type. As well as BAT-26 (a mononucleotide marker commonly unstable in HNPCC tumours), D2S123 (the closest dinucleotide marker to *hMSH2*) and D18S58, they also included a trinucleotide repeat (FABP2), a tetranucleotide repeat (D15S1232) and BAT-17 which comprises a mononucleotide (A) marker in exon 10 of *hMSH6*. As expected, overall instability was low in both colon and endometrial cancers studied, only one endometrial tumour being classed as MSI-high (unstable at > 33% of loci). In addition to the *hMSH6* mutation in this individual, an *hMLH1* frameshift mutation was also discovered that would probably contribute to the increase in MSI. None of the tumours demonstrated instability at either the tri or tetranucleotide loci, the two most frequent markers showing instability being D2S123 and D18S58 (43% of tumours) (Wu *et al.*, 1999). Data from the mononucleotide markers however did not confirm either way whether *hMSH6* mutations would have a specific effect on polymorphic A tracks as had been suggested due to its involvement in repairing single nucleotide mismatches (Drummond *et al.*, 1995). Given the Amsterdam criteria which stipulates HNPCC individuals develop symptoms prior to 50 years of age, Chan *et al.* (1999) compared individuals who had developed colon cancer at either less than 46 years or greater than 46 years of age and found a significant effect of age with microsatellite instability. Ten different loci including D2S123, D3S1067 (the closest dinucleotide marker to *hMLH1*), D18S58 and BAT-26 were analysed and those tumours demonstrating instability at 4 or more loci were designated MSI (microsatellite instability)-high while tumours showing instability at less than 4 loci were designated MSI-low. Of the younger than 46 years of onset age group, 83% were MSI-high compared to 53% of the older than 46 years of onset group (Chan *et al.*, 1999). Altering the age boundary to 36 years further increased the incidence of MSI-high in the younger set compared to the older set.

Other factors that have an effect on microsatellite instability include whether the tumours are metachronous (i.e. metastasised) or synchronous, their position in the colon or whether MMR genes are hypermethylated. Analysis of markers including BAT-26, D2S123 and D18S57 found an increased number of sporadic tumours demonstrated MIN+ if metachronous rather than synchronous (from 14% to 19%), whereas HNPCC tumours displayed the opposite

trend (from 89% to 75%) (Pedroni *et al.*, 1999). This suggests that two different mechanisms of tumour formation are operating in sporadic and HNPCC tumours. Tumours that were in the right side of the colon were also more frequently MIN+ than those positioned in the left or rectal sides (Pedroni *et al.*, 1999), though it is not clear why tumour position would have an effect. Previous studies had already found increased microsatellite instability in *hMLH1* defective tumours compared to tumours defective in other MMR genes (Lamberti *et al.*, 1999, Wu *et al.*, 1999) and with the discovery that methylation of the *hMLH1* promoter can also inactivate the gene, Kuismanen *et al.* (1999) investigated colorectal tumours for an association between MIN+ and methylation status of *hMLH1*. 73% of sporadic colorectal tumours showed alterations in DNA methylation compared to normal colon tissue, and 61% of MIN+ tumours showed the presence of methylation at sites that were not methylated in normal tissue DNA. The opposite was true for MIN- sporadic tumour DNAs in which 53% displayed the absence of methylation at sites that were methylated in normal DNA (Kuismanen *et al.*, 1999). However, when comparing *hMLH1* mutated MIN+ HNPCC tumour DNA to the sporadic MIN+ cases, no bias was found in the methylation pattern at the same sites and so it seems that promoter hypermethylation is a special characteristic of the true sporadic MIN+ subset and not MIN+ cases as a whole (Kuismanen *et al.*, 1999). Investigations into whether other genes were also hypermethylated in the sporadic tumours revealed that a significant proportion (26%) of the MIN- that were hypomethylated at *hMLH1* were also hypomethylated at the calcitonin gene which is a well characterised methylation target in tumours. MIN+ tumours hypermethylated at *hMLH1* were also hypermethylated at calcitonin (Kuismanen *et al.*, 1999) and so it seems that there are different development pathways in MIN+ and MIN- tumours.

The contrasting data available for different types of tumours suggests that different mechanisms are contributing towards tumour formation. HNPCC tumours demonstrate more widespread microsatellite instability, whereas sporadic, RER- colon tumours have elevated mutations in *KRAS* and *p53*. Inherited germline mutations in DNA mismatch repair genes explained the differences between the two types, with new mutations being documented regularly. The latest DNA repair gene to be identified, *MLH3* (Lipkin *et al.*, 2000) may shed more light onto the interactions between these proteins and why mutations in *hMLH1* and *hMSH2* have more severe consequences than mutations in *hPMS1, hPMS2* or *hMSH6*.

1.7.2 Ovarian Cancer

Ovarian cancer accounts for 4% of all cancers in women and 5-10% of these are hereditary. Ovarian cancer can be divided into various subtypes classed as low malignant potential (LMP) or non-invasive, endometrioid, mucinous (endocervical), serous (tubal) and undifferentiated with declining survival rates respectively, and then further subclassified into benign, borderline and malignant according to their degree of proliferation, cytological atypia (i.e. different to the surrounding tissue) and the presence or absence of stromal invasion (Diebold, 1999). LMP, endometrial and mucinous ovarian cancer are usually detected earlier (usually at FIGO - International Federation of Gynecology and Obstetrics- stage I) than serous or undifferentiated (usually at FIGO stage II or III) which probably contributes to the declining survival rates. LMP (borderline) serous tumours develop 10-15 years earlier than invasive serous and do not involve stromal invasion, resulting in a good prognosis.

The inheritance patterns of ovarian cancer can be divided into three subtypes : (a) site-specific in which only ovarian cancers occur (10-15% of inherited ovarian cancer cases), (b) breast-ovarian cancer syndrome in which families show excess breast and ovarian cancer development (65-75% of cases) and (c) ovarian cancer associated with HNPCC type II (10-15% of cases) (Bewtra *et al.*, 1992, Boyd and Rubin, 1997). Various oncogenes and tumour suppressors important to the development of ovarian cancer have been studied, including *p53* (17p13.1), *BRCA-1* (17q21), *BRCA-2* (13q12.3), *ERBB2* (17q21.1), *MYC* (8q24), *KRAS* (12p12.1), *BCL2* (18q21.3) and the *cyclin D1* gene (11q13). The *BRCA-1* gene consists of 22 exons distributed over 100 kbp and putatively codes for a transcription factor that is expressed in rapidly proliferating tissues undergoing differentiation (Miki *et al.*, 1994). The *BRCA-2* gene consists of 26 exons distributed over 70 kbp and possesses sequence and expression pattern similarity to *BRCA-1* (Tavtigian *et al.*, 1996). Mutations in *BRCA-1* and *BRCA-2* are responsible for approximately two thirds of the breast-ovarian cancer syndrome and the range of mutations is so numerous that most are found in only one or two families (Offit, 1996). Germline mutations in *BRCA-1* are responsible for approximately 80% of hereditary ovarian cancers occurring in breast and ovarian cancer syndrome with 89% of these ovarian tumours being serous or undifferentiated (Boyd and Rubin, 1997), whereas the incidence of ovarian cancer in *BRCA-2* families is very low with only late onset development (Takahashi *et al.*, 1996). It is also important to note that ovarian cancer is most likely to occur if the inherited mutation is in the first two-thirds of the *BRCA-1* gene and that various non-pathogenic polymorphisms of the *BRCA-1* gene have been found to be associated with ovarian cancer in families (Janezic *et al.*, 1999) agreeing with an earlier study demonstrating that an increased number of patients had the L871P polymorphism compared to controls (Durocher *et al.*, 1996). The frequency of *BRCA-1* and *BRCA-2* mutations in sporadic ovarian cancers is

low, although deletions of 17q and 13q are frequent in tumours of this type (Jacobs *et al.*, 1993, Kirchweger *et al.*, 1994), suggesting that *BRCA-1* and *BRCA-2* have a role in the development of hereditary cancers and loss of other genes close to them may be responsible for the development of sporadic cancers.

KRAS mutations are more frequent in mucinous (45-75%) than nonmucinous (14-32%) ovarian neoplasms (Enomoto *et al.*, 1991, Pieretti *et al.*, 1995). In fact this is what distinguishes them from other histological types and can be detected in benign, borderline and malignant tumours with increasing frequency (Pieretti *et al.*, 1995, Cuatrecasas *et al.*, 1997). The detection of these *KRAS* mutations in all three subtypes suggests that they are an early event in the development of ovarian mucinous carcinomas. *ERBB2* codes for a cell surface glycoprotein homologous to the epidermal growth factor receptor and amplification or overexpression is common in breast and ovarian carcinomas (Slamon *et al.*, 1989) and associated with a poor prognosis (Berchuck *et al.*, 1990). *p53* is a tumour suppressor gene encoding a nuclear phoshoprotein involved in the regulation of cell division and mutations resulting in prolonged half-life of the protein occur in approximately 60% of tumours (Matias-Guiu and Prat, 1998). *p53* mutations occur more frequently in ovarian carcinomas (Milner *et al.*, 1993, Frank *et al.*, 1994, McManus *et al.*, 1994) than borderline tumours (Wertheim *et al.*, 1994, Kupryjanczyk *et al.*, 1995) and since ovarian cancer is not characteristic of Li Fraumeni families inheriting a germline *p53* mutation and accumulation of the protein does not occur in FIGO stage I, this suggests that *p53* mutations are not early events in ovarian cancer.

BCL2 is an antiapoptotic protein regularly found in non-neoplastic epithelium of the ovary (Henriksen *et al.*, 1995) and hence expression is associated with a good prognosis. In fact *BCL2* expression seems to help counteract the *p53* adverse effects when coexpressed at the same time, resulting in a better prognosis (Diebold *et al.*, 1996). The *cyclin D1* gene is a member of mammalian cyclins divided into five major subclasses (A-E). Cyclins C, D1-3 and E regulate the G1 check point of the cell cycle whereas cyclins A and B1-2 regulate the transition to mitosis (Matias-Guiu and Prat, 1998). Increased expression of *cyclin D1* has been found in 26% of ovarian tumours, most commonly associated with borderline and well-differentiated carcinomas (Worsley *et al.*, 1997).

Each class of ovarian cancer has various levels of genome instability and hence different levels of microsatellite instability. Many studies have focused on invasive ovarian tumours with few benign or LMP cases included solely due to late detection in patients, but it is clear that as the tumours progress to more aggressive stages, MIN+ is less evident and LOH and mutations in oncogenes and tumour suppressors are more frequent (Fujita *et al.*, 1995, King *et al.*, 1995,

Pieretti *et al.*, 1995, Phillips *et al.*, 1996, Tangir *et al.*, 1996, Shih *et al.*, 1998, Wiper *et al.*, 1998). In order to class a tumour as MIN+, two out of at least five microsatellite markers have to demonstrate instability when compared to normal tissue (either blood or non-tumourous) but in many studies only one unstable locus classes the tumour as MIN+. When analysing microsatellite instability, either a small number of microsatellite markers are analysed in large numbers of tumours with varying subtypes, or a broad spectrum of markers are analysed in only one or two subtypes of carcinoma. Shih *et al.* (1998) analysed 69 microsatellite markers distributed over all autosomes including D2S123, D17S5, D17S855, D17S1322, D18S50 and the X chromosome in 31 LMP tumours and 23 cystadenomas and found no microsatellite instability or loss of heterozygosity (LOH - i.e. one allele is lost). Due to reports of frequent LOH at chromosome 17p13 (location of *p53* and other candidate genes *OVCA1, OVCA2* and *HIC-1*) in both early and late-stage tumours (Dodson *et al.*, 1993, Foulkes *et al.*, 1993, Phillips *et al.*, 1993), five chromosome 17p13 markers including one intragenic to *p53* were analysed in early-stage epithelial ovarian cancers (Wiper *et al.*, 1998). Of the 29 tumours analysed, two displayed microsatellite instability at marker D17S695, while the majority had increasing LOH with more advanced stages of tumour progression in this region. However, one of the MIN+ samples may have been from a HNPCC family as the patient's mother died from colon cancer (Wiper *et al.*, 1998). This agreed with an earlier study which discovered MIN+ on chromosome 17 in only one sample out of 57 sporadic ovarian carcinomas classed into various subtypes and at different FIGO stages (Phillips *et al.*, 1996). This tumour came from an individual whose father had died of colon cancer and who was herself diagnosed 20 years earlier than the average age of onset in this study. The tumour type was serous FIGO stage III (Phillips *et al.*, 1996). A second extensive study of chromosome 17 analysed 20 markers distributed on both arms which encompassed *p53* and *BRCA-1* in endometrioid, mucinous and serous ovarian tumours of LMP, low, medium and high grade types (Pieretti *et al.*, 1995). One marker on each of the remaining autosomes and the X chromosome was also analysed for MIN and LOH status as well as *KRAS* activation in the tumour samples. The results demonstrated that increasing severity of ovarian cancer, either in subtype or grade is associated with decreasing levels of MIN+ and increasing LOH (Pieretti *et al.*, 1995). LOH on chromosome 17 is rare in mucinous tumours, whereas increasing loss of sections of this chromosome seem to be associated with advancing stages of serous carcinomas. Mucinous tumours also have significantly higher levels of *KRAS* activating mutations than other types.

Analysis of LOH at 12 markers over 8 chromosomes including 17q21 (D17S1322) and 18q21 (D18S34) confirmed that widespread LOH is a feature of invasive carcinomas whereas borderline tumours have very little LOH, the most commonly affected markers being near the p53 and BRCA-1 loci (Saretzki *et al.*, 1997). Generally, increased LOH is associated with

either advanced FIGO stages or poor differentiation but not survival outcome. Microsatellite markers on chromosome 3 distributed throughout both arms demonstrated a significant difference between borderline and invasive carcinomas in regard to LOH and MIN status (Tangir *et al.*, 1996). Borderline tumours (BOT) which generally have a more favourable prognosis possessed very little LOH (5.6% versus 39%) but greater frequency of MIN+ (28% versus 6.5%) than invasive tumours, however the few borderline tumours displaying MIN+ were classed as serous, whereas the mucinous borderline tumours did not demonstrate MIN+ status (Tangir *et al.*, 1996). Other studies on the same BOT samples showed that MIN was not evident at 46 other markers on chromosomes 6q (Rodabaugh *et al.*, 1995), 9p (Rodabaugh *et al.*, 1995), 17p (Wertheim *et al.*, 1996) and 17q (Tangir *et al.*, 1996). Again, analysis of FIGO stage I-IV carcinomas confirmed the low incidence of MIN in ovarian cancer, and of the tumours that did display MIN+, though only at one locus out of four, they were all classed at FIGO stages I-III (Park *et al.*, 1995).

Comparing ovarian tumour and blood samples from individuals with or without a family history of cancer found no significant difference in the MIN status in either family history or sporadic occurrence, age at diagnosis or histopathology at 3 dinucleotide markers (Arzimanoglou *et al.*, 1996). The rate of MIN+ was low as previously reported, only 11/90 tumours demonstrating instability at least one marker out of three, the most frequently affected being D2S123 (*hMSH2* marker). MIN+ could be divided into three types, either loss of an allele (LOH), additional bands or shifting of the progenitor alleles. Again, there was no correlation between MIN+ type or histopathology, although the only tumour displaying MIN+ at all three markers came from a suspected HNPCC family (Arzimanoglou *et al.*, 1996). King *et al.* (1995) also found low rates of MIN+ in ovarian neoplasms (17%) although only two markers were analysed, a di- and tetranucleotide (*GABARB1*) repeat on chromosomes 5q11-13 and 4p12-13. All seven of the MIN+ tumours were unstable at the tetranucleotide marker and two of these were also unstable at the dinucleotide marker. Patients over the age of 50 years at diagnosis had a slightly higher rate of MIN+ as did those with additional tumours, either synchronous or metachronous and all MIN+ tumours were classed as FIGO stage I and III (King *et al.*, 1995). Fujita *et al.* (1995) also found increasing MIN+ at four dinucleotide markers on chromosomes 5q, 15q, 17p and 18q in the less aggressive endometrioid and mucinous tumours than the serous carcinomas, however MIN+ status was found to be associated with more advanced FIGO stages (although a small number of tumours were analysed) which is contrary to the previous studies detailed above. *hMSH2* mutations were associated with three of the seven MIN+ tumours, however the extent of MIN+ in these samples was not different to the remaining four samples (Fujita *et al.*, 1995).

Comparing LOH, MIN status and mutations in oncogenes such as *KRAS* suggests there are at least two different mechanisms of tumour progression in ovarian cancer. In fact, the most common genetic change in ovarian cancers seems to be loss of chromosome 17, followed by LOH on other chromosomes, mutations in genes such as *KRAS* and then mismatch repair mutations giving rise to MIN+ (Pieretti *et al.*, 1995). Although the rate of MIN+ in ovarian cancer is low, it occurs more frequently in the early stage tumours than the late stages. This MIN+ status is relatively focal and not widespread in the genome, whereas LOH is more frequent in late stage tumours, especially on chromosome 17 and more widespread. More aggressive invasive epithelial ovarian cancers have increased mutations of oncogenes such as *KRAS* and *p53*, but some types of tumour, such as mucinous have elevated *KRAS* mutations that may define their type.

Disease/Locus	Repeat	Allele Sizes	Location	Reference
CTG18.1	CTG	Normal : 11-2000	Putative intron 3 of *SEF2-1* gene at 18q21.1	Breschel *et al* 1997
Dentatorubral-pallidoluysian atrophy (DRPLA)	CAG	Normal : 8-35 Mutation : 49-75	exon 5 of *atrophin-1* gene at 12p13.3	Koide *et al* 1994 Nagafuchi *et al* 1994
Expanded Repeat Domain CAG/CTG 1 (ERDA1)	CAG/CTG	Normal : 7-92	17q21.3	Nakamoto *et al* 1997 Ikeuchi *et al* 1998 Deka *et al* 1999
Fragile X syndrome (FRAXA)	CGG	Normal : 6-54 Premutation : 50-200 Mutation : 200-1000	5' UTR exon 1 of *FMR-1* gene at Xq27.3	Fu *et al* 1991 Reyniers *et al* 1993
FRAXE (mental retardation)	GCC	Normal : 6-25 Mutation : 115-850	5' UTR exon 1 of *FMR-2* gene at Xq28	Flynn *et al* 1993 Knight *et al* 1993
FRAXF	CCG	Normal : 6-30 Mutation : 300->1000	X chromosome	Parrish *et al* 1994
Jacobsen's syndrome (FRA11B)	CCG	Normal : 8-50 Mutation : 100->1000	5' UTR of *CBL2* oncogene at chromosome 11	Jones *et al* 1994 Jones *et al* 1995
FRA16A	CGG	Normal : 17-50 Mutation : 1000->1900	Chromosome 16	Nancarrow *et al* 1994 Nancarrow *et al* 1995
Friedreich's Ataxia (FRDA)	GAA	Normal : 7-22 Mutation : 200-900	Intron 1 of *frataxin* gene at 9q13	Campuzano *et al* 1996

Disease/Locus	Repeat	Allele Sizes	Location	Reference
Huntington's Disease (HD)	CAG	Normal : 9-39 Mutation : 36-121	Exon 1 of *huntingtin* gene at 4p16.3	Huntington's Disease Collaborative Research Group 1993 Zuhlke *et al* 1993
Machado Joseph's Disease (MJD/SCA3)	CAG	Normal : 14-34 Mutation : 61-84	C-terminal coding region of *ataxin-3* gene at 14q32.1	Kawaguchi *et al* 1994
Myotonic dystrophy (DM-1)	CTG	Normal : 5-39 Mutation : 50-2000	3' UTR exon 15 of *DMPK* gene at 19q13.3	Aslanidis *et al* 1992 Brook *et al* 1992 Buxton *et al* 1992 Fu *et al* 1992 Harley *et al* 1992 Mahadevan *et al* 1992
Spinal and bulbar muscular atrophy (SBMA)	CAG	Normal : 17-42 Mutation : 40-52	Exon 1 of *androgen receptor* gene at Xq11.2-q12	La Spada *et al* 1991
Spinocerebellar ataxia type 1 (SCA1)	CAG	Normal : 19-36 Mutation : 43-81	*Ataxin-1* gene at 6p24	Orr *et al* 1993
Spinocerebellar ataxia type 2 (SCA2)	CAG	Normal : 15-29 Expanded : 35-59	*Ataxin-2* gene at chromosome 12	Imbert *et al* 1996 Pulst *et al* 1996 Sanpei *et al* 1996

Disease/Locus	Repeat	Allele Sizes	Location	Reference
Spinocerebellar ataxia type 6 (SCA6)	CAG	Normal : 4-16 Mutation : 17-27	*α voltage-dependent Ca channel subunit* gene at 19p13	Zhuchenko *et al* 1997
Spinocerebellar ataxia type 7 (SCA7)	CAG	Normal : 7-16 Mutation : 41-306	coding region of *ataxin-7* gene at 3p12-13	David *et al* 1997 Benton *et al* 1998
Spinocerebellar ataxia type 8 (SCA8)	CTG	Normal : 13-88 Mutation : 107-127	13q21	Koob *et al* 1999
Spinocerebellar ataxia type 12 (SCA12)	CAG	Normal : 7-28 Mutation : 66-78	5' UTR of *PR55β* gene	Holmes *et al* 1999

Table 1.1. Allele sizes and gene locations of triplet repeats. Allele sizes and the gene locations of human triplet repeats. Numbers in allele sizes depict repeats. Where known, the gene is named.

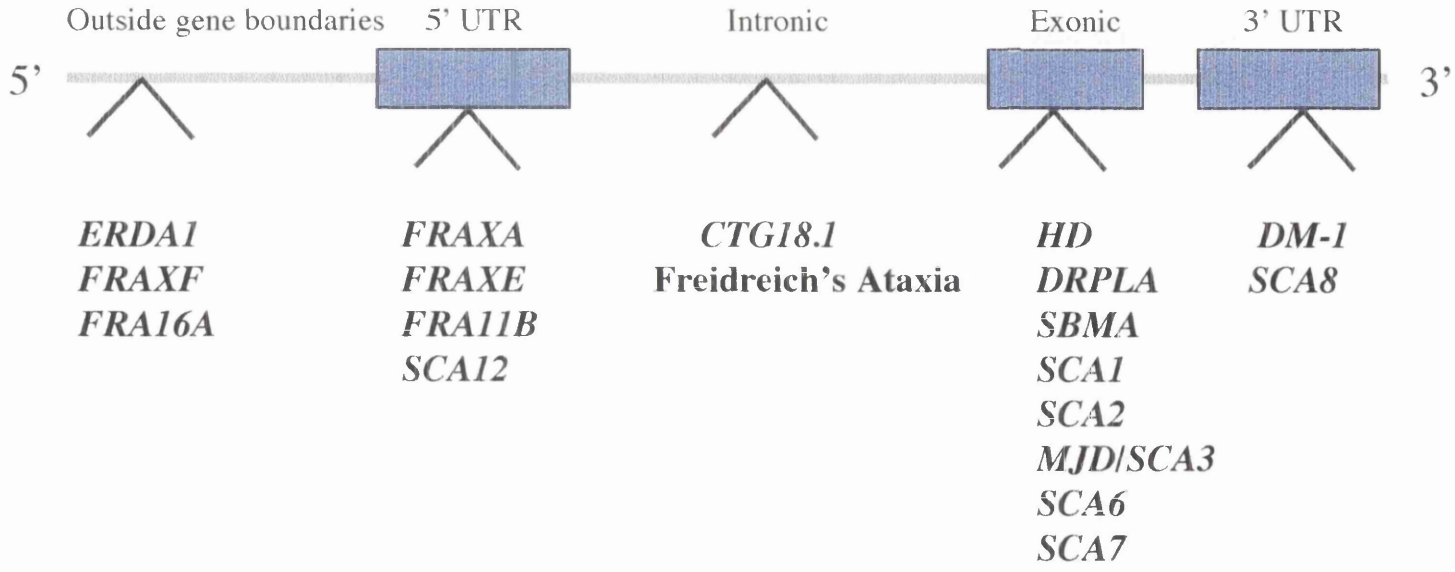

Figure 1.1 Representation of the location of expandable trinucleotide repeats in humans. Displayed is an imaginary gene containing trinucleotide repeat sequences. The expandable triplets discovered are predominatly located within a gene, *ERDA1*,*FRAXF* and *FRA16A* being the exceptions. Repeats in exons are thought to confer a gain of function and the *DM-1* repeat is thought to disrupt the putative *Six5* promoter downstream.

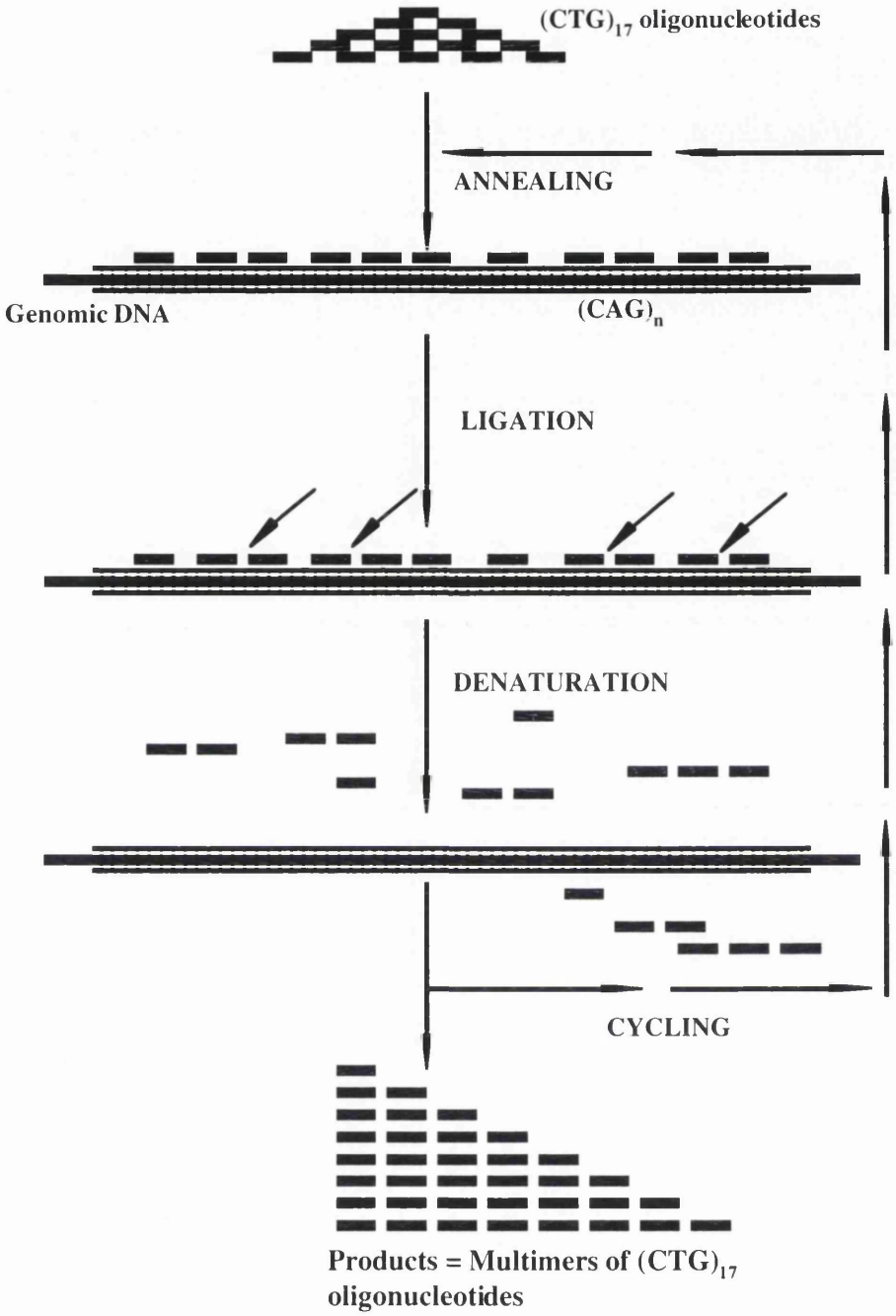

Figure 1.2 Schematic representation of the Repeat Expansion Detection Method (RED). An oligo comprised of multimers of the repeat of interest is annealed to denatured genomic DNA and hybridised to the target at a temperature near its melting point. Ligase then binds adjacent oligos, producing a mixed population of single stranded DNA molecules which can be separated by agarose gel electrophoresis.

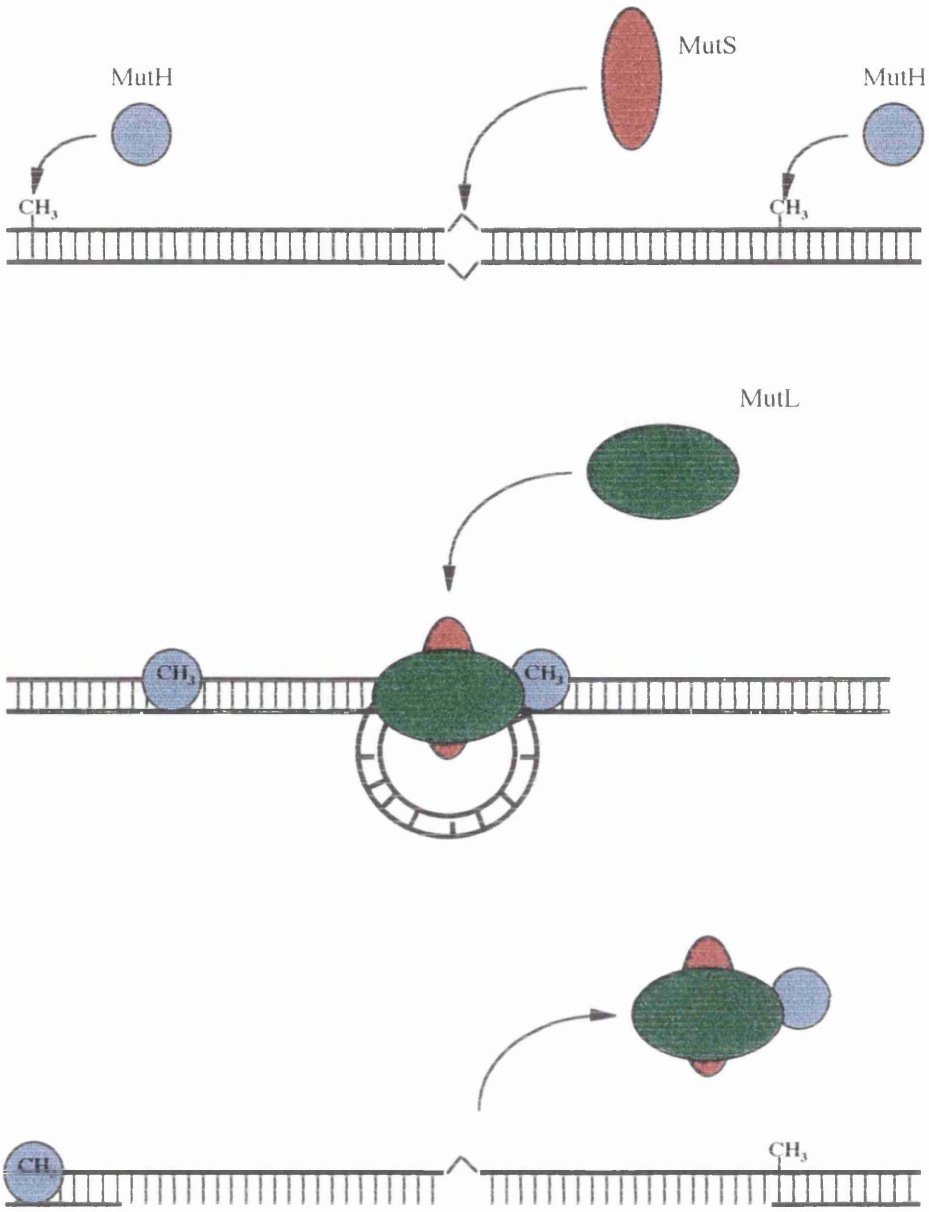

Figure 1.3. Schematic representation of the process of DNA mismatch repair in *E.coli*. Within the repairosome, it is known that MutS recognises and binds to the mismatch in the DNA, with MutH binding to hemimethylated GATC sites on the incorrect strand. MutL is then recruited by MutS, inactivating MutH. MutU (UvrD helicase) acts in conjunction with exonucleases (RecJ and ExoI) to excise the unmethylated strand and the new strand is now synthesised in place of the old, incorrect strand by DNA polymerase III, single-strand DNA-binding protein and DNA ligase.

Chapter 2

Materials and Methods

2.1 Preparation of DNA from tissues

Semen samples were donated by patients attending the Infertility Clinic Unit of the Western Infirmary, Glasgow and subsequently supplied by Dr. Mena Behan. Samples were refrigerated following donation, acquired within 24 hours and frozen at -20°C upon receipt. Sperm counts ranged from 1->500 millions/ml.

2.1.1 Crude preparation of DNA from human tissue

In order to approximate triplet repeat allele lengths at the *CTG18.1* and *ERDA1* loci, 100µl of semen or blood was resuspended in 900µl of 1 x SSC and the cells pelleted by centrifugation at 4,000 rpm for 5 minutes in a bench top centrifuge at room temperature. The supernatant was decanted. 100µl of lysis buffer and 15µl of 20 mg/ml proteinase K was added to the pelleted cells and the mixture incubated overnight at 60°C. Following vigorous vortexing, cellular debris was pelleted at 4,000 rpm for 5 minutes in a bench top centrifuge at room temperature. A small amount of DNA from the supernatant (approx 1µl) was made to 50µl with distilled water in a screw top Eppendorf tube resulting in a final DNA concentration of approximately 5ng/µl. The proteinase K was inactivated by incubating the diluted DNA at 95°C for 5 minutes. The samples were stored at -20°C until required.

2.1.2 Preparation of DNA from semen samples

To directly observe mutations in the male germline cells, good quality DNA needed to be isolated from sperm. The following protocol selectively enriches sperm DNA from the semen samples in preference to any somatic cell contamination that may be present. The protocol is adapted from the Nucleon Biosciences BACC III genomic DNA extraction kit blood and cell culture protocol.

Selected frozen semen samples were thawed on ice and 500µl was resuspended in 2ml of 1 x SSC. The cells were pelleted by centrifugation at 3,000 rpm for 2 minutes in a bench top centrifuge at room temperature, the supernatant discarded and the pellet resuspended once again in 2ml 1 x SSC. In order to remove any somatic cell contamination, 200µl of 10% SDS was added, the samples inverted several times and the intact sperm cells pelleted at 3,000 rpm for 2 minutes at room temperature. 500µl of the supernatant containing somatic cell DNA was aliquoted into a screw top Eppendorf tube and the remaining supernatant discarded. To prepare the sperm DNA, the pellet was resuspended in 2ml reagent B (Nucleon Biosciences) with β-mercaptoethanol and incubated at 37°C in a water bath for 1 hour. 500µl of 5M sodium perchlorate was added, the tubes placed on a blood tube rotator (Stuart Scientific) and spun for 15 minutes at room temperature. The samples were then incubated at 65°C while shaking. 2ml of chloroform was added and spun on the blood tube rotator for a further 10 minutes. The samples were spun at 800 g for 1 minute, 300µl of resin (Nucleon Biosciences) added, mixed by inversion and centrifuged at 1400 g for 3 minutes. 2ml of the upper layer was transferred to a fresh 15ml tube, avoiding the resin. 2.5 x volume 100% ethanol was added and the tubes inverted several times to precipitate the DNA. Following a pulse spin at 4,000 rpm in a bench top centrifuge for 30 seconds, the ethanol was removed and the pellet dissolved in 250µl TE. The dissolved sperm DNA was transferred to a fresh 1.5ml screw top Eppendorf and stored at -20°C.

To purify the somatic DNA from the SDS sensitive fraction, 500µl of buffered phenol was added and the two layers mixed well. The samples were spun at 14,000 rpm for 10 minutes in a bench top centrifuge and the top layer (approximately 500µl) transferred to a fresh 1.5ml screw top Eppendorf. Equal volume of phenol:chloroform:isoamylalcohol (25:24:1) was added and the two layers mixed well again. The samples were spun at 14,000 rpm for 10 minutes in a bench top centrifuge and the upper layer transferred to a fresh 1.5ml screw top Eppendorf. 0.1 x volume of 2M Sodium Acetate and 2.5 x volume 100% ethanol were added and the tubes inverted several times to precipitate the somatic DNA. Following a pulse spin at 14,000 rpm for 30 seconds, the ethanol was removed and the pellet dissolved in 50µl TE.

2.1.3 Preparation of DNA from human tissue

In order to prepare DNA from a fixed tissue sample, the paraffin has to be removed. The following protocol was adapted from the Nucleon Biosciences BACC III genomic DNA extraction kit blood and cell culture protocol.

A $2cm^3$ block of paraffin embedded, formalin-fixed tumour tissue was thinly sliced and mixed in 20ml of xylene in a glass bijou on a blood tube rotator for 1 hour. The xylene was removed and the tissue placed in a 1.5ml screw top Eppendorf containing 550µl lysis buffer and 15µl 20mg/ml proteinase K. The sample was incubated overnight at 60°C in a water bath.

Frozen tissue was thawed at room temperature and 10-20mg minced. The tissue was placed in a 1.5ml screw top Eppendorf containing 550µl lysis buffer and 15µl 20mg/ml proteinase K and incubated overnight at 60°C in a water bath. All samples were centrifuged at 4,000 rpm for 1 minute in a bench top centrifuge and 250µl of the supernatant transferred to a fresh 1.5ml screw top Eppendorf. 100µl of de-ionised water was added and the samples vortexed for 5 seconds. 150µl of 5M sodium perchlorate was added and the tubes inverted at least seven times. 600µl of chloroform was added and the tubes inverted seven times. 150µl of resin was added and the tubes rotary mixed for 5 minutes. The samples were centrifuged at 500 g for 2 minutes. The upper layer was transferred to a fresh 1.5ml screw top Eppendorf, avoiding the resin interface. 1.2ml of 100% ethanol was added and the tubes inverted several times until the DNA precipitated. The samples were centrifuged at 14,000 rpm for 5 minutes in a bench top centrifuge and the supernatant discarded. 1ml of cold 80% ethanol was added and the tubes inverted several times. The samples were centrifuged at 14,000 rpm for 5 minutes in a bench top centrifuge and the supernatant discarded. The DNA pellets were dissolved in 100-200µl TE and heated at 60°C for 1 hour. DNA was stored at -20°C.

2.2 PCR Allele Typing

2.2.1 *CTG18.1*

Samples prepared in section 2.1.1 and 2.1.3 were thawed at room temperature. 5ng of DNA was amplified in 10µl reactions using 1x PCR buffer, 1µM forward primer SEF2-C, 1µM reverse primer SEF2-BR and 0.05 U/µl Taq Polymerase (Amplitaq, Perkin Elmer) in 0.2ml thin-walled tubes. Reactions were cycled through 30 rounds of 96°C (45 seconds), 67°C (45 seconds) and 70°C (3 minutes) and one further cycle of 67°C (1 minute) and 70°C (10 minutes) in a Biometra UNO II 96 thermocycler. 3µl of Orange G loading dye was added and the reaction products separated by electrophoresis through a 2% NuSieve GTG Low melting temperature agarose (Flowgen)/1% Agarose (BM) 20cm gel in 0.5 x TBE buffer at 150 volts for 90 minutes. 600ng of 1kbp Ladder (GibcoBRL) was electrophoresed down each side of the gels. PCR products were visualised on a UV transilluminator (UVP Ltd), captured with a UVP digital camera and photographs printed on Kodak thermal paper.

2.2.2 *CTG18.1* fluorescent typing

Samples prepared in section 2.1.1 were thawed at room temperature. 5ng of DNA was amplified in 10µl reactions using 1x PCR buffer, 1µM fluorescent forward primer SEF2-C-F, 1µM reverse primer SEF2-BR and 0.05 U/µl Taq Polymerase (Amplitaq, Perkin Elmer) in 0.2ml thin-walled tubes. Reactions were cycled through 28 rounds of 96°C (45 seconds), 67°C (45 seconds) and 70°C (3 minutes) and one further cycle of 67°C (1 minute) and 70°C (10 minutes) in a Biometra UNO II 96 thermocycler. 2µl of the reaction products was transferred to a fresh 0.2ml thin-walled tube, removing any excess oil from the tip before transferring. 1ul of GeneScan standard mix was added in order to mix the molecular weight standard with the PCR products and the samples denatured at 90°C for 2 minutes in a Biometra UNO II 96 thermocycler. Following a brief spin in a bench top centrifuge, samples were electrophoresed through a 4.75% urea polyacrylamide gel in a 373XL automated sequencing machine (Perkin Elmer) for 5-12 hours on Filter Set A at 40 watts, depending on the size of the alleles. The resulting Gel File was analysed with GeneScan software version 3.1 (Perkin Elmer).

2.2.3 *ERDA1*

Samples prepared in section 2.1.1 and 2.1.3 were thawed at room temperature and underwent typing PCR at the *ERDA1* locus in an identical way to the *CTG18.1* locus. Primers ERDA-C and ERDA-DR were substituted for primers SEF2-C and SEF2-BR and the annealing temperature was 62°C.

2.2.4 *ERDA1* fluorescent typing

Samples prepared in section 2.1.1 and 2.1.3 were thawed at room temperature and underwent fluorescent typing PCR at the *ERDA1* locus in an identical way to the *CTG18.1* locus. Primers ERDA-A-F and ERDA-DR were substituted for primers SEF2-C-F and SEF2-BR and the annealing temperature was 62°C.

2.2.5 *ERDA1* A/C polymorphism typing

Chimpanzee DNA and samples prepared in section 2.1.1 were thawed at room temperature and underwent typing PCR at the *ERDA1* locus in an identical way to Section 2.2.3. Primers ERDA-ERG or ERDA-ERT were separately substituted for primer ERDA-DR in identical reactions with a 62°C annealing temperature.

2.3 SP-PCR

DNA samples from sperm as prepared in section 2.1.2 and from blood, tumours and cell lines supplied by Dr. Tetsuo Ashizawa, Dr. Robert Brown, Prof Patricia Jacobs and Dr. Anne Murray, Dr. Marvin Meistrich, Dr. Michael Siciliano, Dr. Margot Thomas and Centre d'Etude du Polymorphisme Humain (CEPH) were digested with *Hin*dIII and diluted in 10mM Tris-HCl pH 7.5, 1mM EDTA and 0.1µM carrier forward primer (to combat reduction of genomic DNA concentration during storage). 12pg - 50 ng of *Hin*dIII digested DNA was amplified in 7µl reactions using 1 x PCR buffer, 0.2µM forward primer, 0.2µM reverse primer (Table 2.1) and 0.05 U/µl of Taq Polymerase (Amplitaq, Perkin Elmer). Reactions were cycled through 28 rounds of 96°C (45 seconds), 55-68°C (45 seconds) and 70°C (3 minutes) and one further cycle of 55-68°C (1 minute) and 70°C (10 minutes) in a Biometra UNO II 96 thermocycler. 4.2µl of the reaction products were electrophoresed through a 1.75% 20cm agarose gel in 0.5 x TBE buffer at 90 volts for 17 hours. 1µg each of Amplisize and 2.5kb ladder molecular weight markers (BioRad) respectively were electrophoresed along the side and down the middle of each gel. Gels were Southern blotted overnight and hybridised with 20ng of a CTG repeat probe and 6.7ng of each of the Amplisize and 2.5kb ladder (Section 2.5). Resultant alleles were detected by autoradiography (16h-7days) with exposure to an intensifying screen at room temperature.

2.4 Single molecule fluorescent PCR

0.7µl of the single molecule reaction products from Section 2.3 were diluted 1:100 in de-ionised water and 1µl re-amplified in 10µl reactions using 1 x PCR Buffer, 0.2µM forward FAM fluorescent primer (Oswell), 0.2µM reverse primer and 0.05 U/µl of Amplitaq. Reactions were cycled through 19 rounds of 96°C (45 seconds), 57 °C (*ERDA1*) or 62°C (*CTG18.1*) (45 seconds) and 70°C (3 minutes) and a chase of 57 °C (*ERDA1*) or 62°C (*CTG18.1*) (1 minute) and 70°C (10 minutes) in a Biometra UNO II 96 thermocycler. Reaction products were electrophoresed on a 4.75% polyacrylamide gel as previously described in Section 2.2.2 and analysed with Genescan 3.1 software.

Primer Name	**Sequence (5'-3')**
DM-A (Forward)	CAGTTCACAACCGCTCCGAGC
DM-BR (Reverse)	CGTGGAGGATGGAACACGGAC
DM-C (Forward)	AACGGGGCTCGAAGGGTCCT
DM-C-F (Fluorescent Forward)	**FAM-SPACER**-AACGGGGCTCGAAGGGTCCT
ERDA-A (Forward)	CTATCAATGGCTCCAAGTAGCAC
ERDA-A-F (Fluorescent Forward)	**FAM-SPACER**-CTATCAATGGCTCCAAGTAGCAC
ERDA-BR (Reverse)	ATCCTTTTGCAACTAGCGGTGAC
ERDA-C (Forward)	GGAAGGCATCTTCAAAATGGATTG
ERDA-DR (Reverse)	GAAAAATAAGGTGGAAGGAAGGTC
ERDA-ERG (Reverse)	ATAAGGTGGAAGGAAGGTCTTTTG
ERDA-ERT (Forward)	ATAAGGTGGAAGGAAGGTCTTTTT
SBMA-F (Forward)	GCCTGTTGAACTCTTCTGAGC
SBMA-R (Reverse)	GAGGAACAGCAACCTTCACAGC
SCA1-F (Forward)	CAACATGGGCAGTCTGAG
SCA1-R (Reverse)	CTCAGACTGCCCATGTTG
SEF2-BR (Reverse)	AGGAACGAATGGAGAAAGTGCAAC
SEF2-C (Forward)	AATCCAAACCGCCTTCCAAGTG
SEF2-C-F (Fluorescent Forward)	**FAM-SPACER**-AATCCAAACCGCCTTCCAAGTG

Table 2.1 Sequences of PCR primers. Depicted are the primer sequences used in typing, fluorescent and SP-PCR. FAM is the blue fluorescent tag 6-carboxyfluorescein. The spacer is comprised of a random six nucleotide sequence to facilitate the fluorescent primer binding.

2.5 Southern blot hybridisation

Agarose gels were submerged in Depurinating Solution for 10 minutes, rinsed with de-ionised water then submerged in Denaturing Solution for 30 minutes. Gels were rinsed with de-ionised water then submerged in Neutralising Solution for 30 minutes. Hybond-N nylon membrane (Amersham) equilibrated in Neutralising Solution was placed onto the gel which was laid on a flat surface. Squash blotting was set up by stacking equilibrated Whatman 3MM paper, dry absorbant paper, a glass plate and a 2Kg weight sequentially on top of the gel. Transfer of DNA from the gel then took place overnight at room temperature. DNA was fixed to the membrane by baking at 80°C for 3 hours, followed by exposure to ultraviolet light (UVP transilluminator) for 20 seconds.

20ng of a $(CTG)_{56}$ repeat PCR product and 6.7ng of each of the Amplisize (BioRad) and 2.5kb ladder (BioRad) molecular weight markers in 32µl final volume were denatured at 100°C for 5 minutes, quenched on ice and 10µl OLB (Oligo Labelling Buffer), 2µl 1mg/ml BSA, 5µl α^{32}PdCTP (3000 Ci/mmole) and 2U Klenow (GibcoBRL) added. The labelling reaction was incubated at 37°C for 1 hour. A 1ml Sephadex G-50 column was prepared in a 15ml falcon tube with polymer wool (Interpet) (acting as a plug) and 4% (w/v) Sephadex G-50 in TE. The sephadex was compacted by centrifugation at 2,000 rpm for 2 minutes. The probe was added to the column and centrifuged a second time at 2,000 rpm for 2 minutes to elute the probe into a 1.5ml Eppendorf tube. Purified probe was denatured at 100°C for 5 minutes immediately prior to use.

Membranes were placed in a hybridisation bottle with 30ml of Hybridisation Solution and rotated at 65°C for a minimum of 1 hour. The probe was incubated with the membrane at 65°C overnight in 3ml fresh Hybridisation Solution in a Biometra rotating hybridisation oven. Filters were washed twice at high stringency using Wash Solution at 65°C rotating. The filters were dried at 80°C and exposed to X-Ray film (GRI) in the presence of an intensifying screen (GRI) at room temperature. Autoradiographs were developed initially after overnight exposure and subsequently after 16 hours to 7 days.

2.6 Reagents

1x SSC

NaCl	0.15M
Tri-Sodium Citrate	0.15M

1x TBE buffer

Tris	0.09M
Boric Acid	0.09M
EDTA pH 8.0	0.002M

1x PCR Buffer

Tris HCl pH 8.8	45mM
$(NH_4)_2SO_4$	11mM
$MgCl_2$	4.5mM
β-Mercaptoethanol	6.7mM
EDTA	4.4μM
Deoxyadenosine triphosphate	1mM
Deoxycytidine triphosphate	1mM
Deoxyguanosine triphosphate	1mM
Deoxythymidine triphosphate	1mM
BSA	113 μg/ml

50x Te

Tris HCl pH 7.5	0.5M
EDTA pH 8.0	5mM

Blue Loading Dye

Glycerol	50% (v/v)
5 x TBE	50% (v/v)
Bromophenol Blue	2μg/ml
Xylene Cyanol	2μg/ml

Denaturing Solution

NaCl	0.5M
NaOH	0.4M

Depurinating Solution

HCl	0.25M

GeneScan standard mix

Formamide	
Blue Loading Dye (Perkin Elmer)	
Flourescent MW Standard GS2500 ROX (Perkin Elmer)	4.64:1:1.08 (v/v)

Hybridisation Solution

EDTA	1mM
SDS	7% (w/v)
Sodium Phosphate pH7.2	0.5M

Lysis Buffer

Tris HCl pH 8.0	50mM
EDTA pH 8.0	100mM
SDS	0.5% (w/v)

Neutralising Solution

Tris	0.5M
NaCl	3.0M
pH to 7.5 with concentrated HCl	

Oligo Labelling Buffer (OLB)

TE	30% (v/v)
Random Hexamers	30 OD units/ml
Hepes pH 6.6	1M
β-Mercaptoethanol	0.35% (v/v)
Deoxyadenosine triphosphate	0.1mM
Deoxycytidine triphosphate	0.1mM
Deoxyguanosine triphosphate	0.1mM
Deoxythymidine triphosphate	0.1mM
Tris-HCl pH 8.0	0.24M
$MgCl_2$	2.4mM

Orange G Loading Dye

Orange G Powder	0.6µg/ml
Glycerol	50% (v/v)
de-ionised H_2O	50% (v/v)

TE

Tris HCl pH 7.5	10mM
EDTA pH 8.0	1mM

Wash Solution

SDS	0.2% (w/v)
SSC	0.2x (v/v)

2.7 Abbreviations

%GC	Percentage of Guanosine and Cytidine nucleotides in a DNA sequence
bp	Base pairs
BSA	Bovine Serum Albumin
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
FAM	6-carboxyfluorescein
kbp	Kilobase pairs
Ln	Natural Log
LOH	Loss of heterozygosity
MMR	Mismatch repair
MSI	Microsatellite instability
NIX	Nucleotide Identification package
No.	Number
PCR	Polymerase chain reaction
ROX	6-carboxy-x-rhodamine
rpm	Revolutions per minute
Rpts	Repeats
SDS	Sodium Dodecyl Sulfate
Tris	Tris(hydroxymethyl)aminomethane
UTR	Untranslated Region

Chapter 3

Allele length diversity at the *CTG18.1* and *ERDA1* loci in the general population

3.1 Introduction

Since the discovery of dynamic mutations giving rise to inherited diseases, much work has focused on the somatic and germline behaviour of these trinucleotide repeats which has defined certain differences dependent on the sequence and its purity, location in respect to the gene and the length of the alleles. Expandable trinucleotide repeats give rise to diseases such as HD, SBMA and the spinocerebellar ataxias and in cases such as *DM-1*, have been documented to attain greater than 2,000 repeats in length in congenitally affected individuals. Recent collated data demonstrates that the trinucleotide diseases do have differing rates of mutability, with some evidence that high GC content in the flanking sequence and the proximity of CpG islands could be a major contributor (Brock *et al.*, 1999). Recently two novel, expandable trinucleotide repeats were discovered on chromosomes 17q21.3 and 18q21.1, termed *ERDA1* and *CTG18.1*. These loci are not yet associated with a disease but have a greater frequency of expanded alleles than any of the previously characterised loci. This makes them ideal candidates on which to observe how repeat loci can be affected by factors such as flanking sequence, inherited mutations in MMR genes and external environmental influences.

3.2 Results

Semen samples collected from an infertility clinic and blood samples from *DM-1* or HNPCC patients were lysed and PCR performed across the polymorphic repeats at the *CTG18.1* and *ERDA1* loci. Crude DNA was prepared by a simple lysis procedure (Section 2.1.1) and PCR was carried out using primers SEF2-C/BR (Section 2.2.1) or ERDA1-C/DR (Section 2.2.3). The PCR products were electrophoresed through a 2% NuSieve GTG Low melting temperature agarose/1% Agarose TBE gel and visualised on an UV transilluminator (Figures 3.1 and 3.2). Shown in Figures 3.1 and 3.2 are examples of typing in either blood or sperm DNA. 628 and 616 chromosomes were typed at the *CTG18.1* and *ERDA1* loci respectively in the semen samples in order to select candidates for SP-PCR analysis and to provide allele distributions for each locus. Products ranging from 10-91 repeats and 7-82 repeats were observed at the *CTG18.1* and *ERDA1* loci respectively. Samples where only one allele was

detected were checked for homozygosity by PCR Southern blot (not shown). No samples demonstrated a second allele at either locus.

The agarose gels gave an approximation of allele sizes. Consequently, PCR was repeated at both loci utilising a fluorescently tagged forward primer (Sections 2.2.2 and 2.2.4) and the products electrophoresed through a 4.75% denaturing urea polyacrylamide gel in a 373XL automated sequencing machine. Sample data was analysed using GeneScan 3.1 software following visualisation (Figures 3.3 and 3.4). The GeneScan software is capable of sizing PCR products to within a base pair due to internal fluorescent molecular weight size standards run with the PCR product in each lane of the urea polyacrylamide gel. The fluorescent size standard is detected as Red, whereas PCR products can be detected as Blue, Green or Yellow depending on the fluorescent tag. In this case, products were labelled with a FAM tag, detected as Blue on Filter Set A. The peak with the greatest area was used to type each chromosome (Figure 3.5) and the total frequencies of each allele plotted for each locus (Figures 3.6 and 3.7).

3.3 Discussion

The allele range of 10-91 repeats at the *CTG18.1* locus fits well with published data by (Breschel *et al.*, 1997). The most common alleles are 11 repeats (34.6%), 14 repeats (16.1%) and 17 repeats (14%) and an expansion (>50 repeats) frequency of 2.9%. The allele range at the *ERDA1* locus also correlates well with published data from Nakamoto *et al.*, (1997), Ikeuchi *et al.*, (1998) and Deka *et al.*, (1999). The most common alleles are 10 repeats (31.5%), 21 repeats (17.1%) and 22 repeats (12.4%) and an expanded allele frequency of 15.1%. Both loci seem to display trimodal distributions. *CTG18.1* demonstrates groups of alleles from 10-18, 19-41 and 60-91 repeats, with no detectable alleles between 41 and 60 repeats. *ERDA1* displays groups of alleles from 7-11, 13-29 and 49-82 repeats with only two intermediate alleles of 36 and 44 repeats detected.

The distribution of alleles at these two loci would seem to infer that the *ERDA1* locus is more stable than the *CTG18.1* locus. Comparison of the frequencies of similarly sized alleles would suggest that the original allele sizes of 11 and 10 repeats at *CTG18.1* and *ERDA1* respectively, underwent rare expansion events to larger alleles of 14 and 17, and 17 and 21 repeats. It is possible that these larger alleles have undergone more common mutations of small length changes, likely to be due to their increase in size and contributed to the spread of alleles in the normal range. Extrapolating on this point it is also possible that the pool of alleles from 50-100 repeats at either locus derived from a rare expansion originating within its

individual 20-40 repeat range. *ERDA1* has an increased frequency of alleles in the 50-90 repeat range which suggests that these alleles are more stable and maintain their frequency in the human population. This is supported by the detection of extremely large *CTG18.1* alleles (Breschel *et al.*, 1997) suggesting that this locus has the potential to produce huge expansions that have so far not been detected at the *ERDA1* locus.

The increase in extra bands detected in the PCR products with increasing repeat length (normally seen as a smear) indicates there may also be instability in the DNA typed. Using small quantities of DNA in PCR reactions at other loci such as *DM-1* and *SCA7* has demonstrated that this smear can be resolved into multiple bands. PCR products from the *ERDA1* locus possess tighter bands than equivalent sized alleles at *CTG18.1* on both agarose and polyacrylamide gels, further suggesting that the *CTG18.1* locus could be more unstable than the *ERDA1* locus.

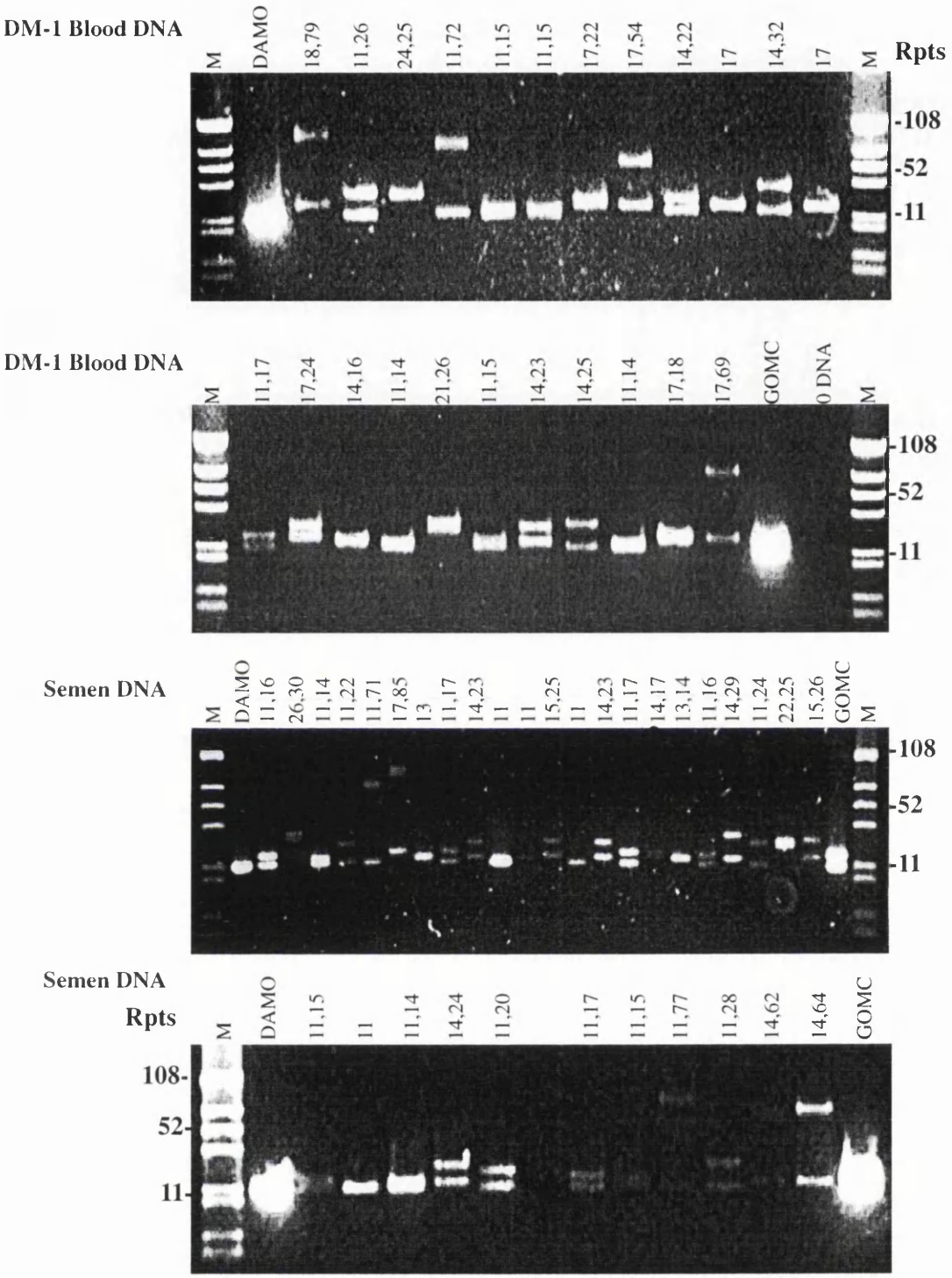

Figure 3.1 Allele typing at the *CTG18.1* locus. Each typing reaction was carried out with primers SEF2-C/BR. Control PCR on genomic DNA DAMO (11,11) and GOMC (11,17) was also carried out for approximation of allele sizes and on a zero DNA control. PCR products were electrophoresed through a 2% NuSieve/1% Agarose gel and visualised on a UV transilluminator. Allele sizes in repeats are depicted for each sample and 1Kb ladder loaded in the marker lane (M).

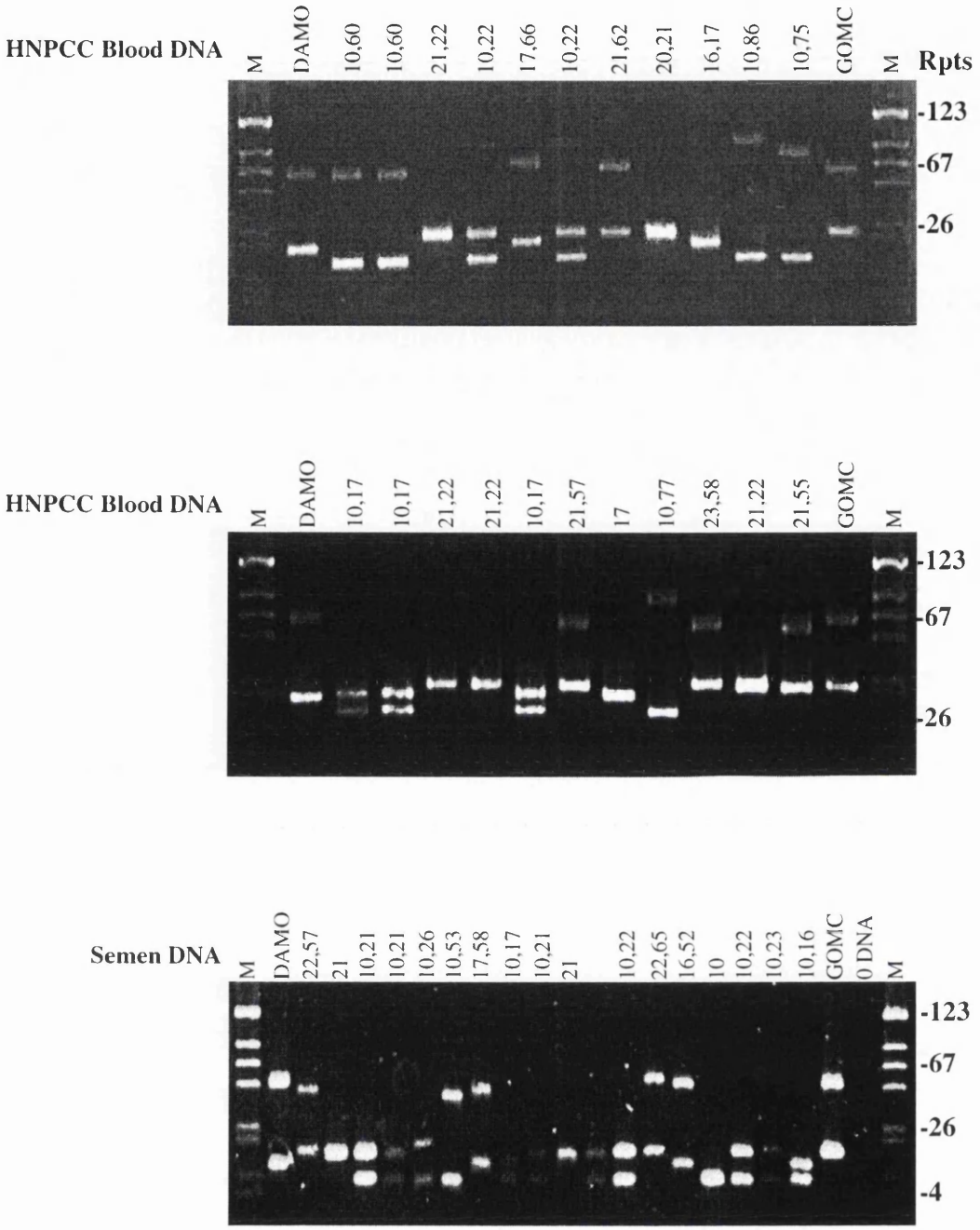

Figure 3.2 Allele typing at the *ERDA1* locus. Each typing reaction was carried out with primers ERDA-C/DR. Control PCR on genomic DNA DAMO (19,62) and GOMC (22,60) was also carried out for approximation of allele sizes and on a zero DNA control. PCR products were electrophoresed through a 2% NuSieve/1% Agarose gel and visualised on a UV transilluminator. Allele sizes in repeats are depicted for each sample and 1Kb ladder loaded in the marker lane (M).

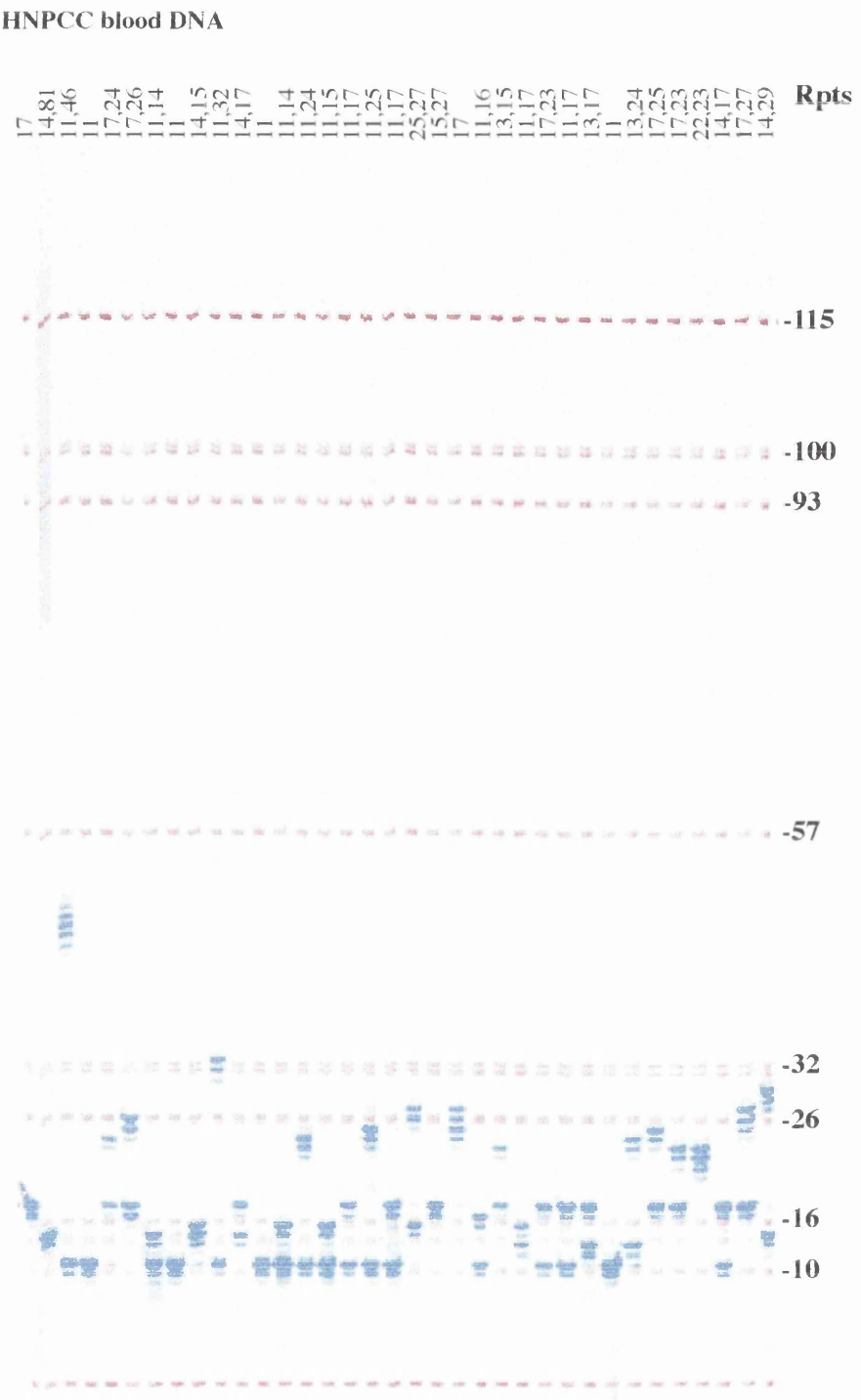

Figure 3.3 Fluorescent allele typing at the *CTG18.1* locus. Each typing reaction was carried out with primers SEF2-C-F/BR. Control PCR on genomic DNA DAMO (11,11) and GOMC (11,17) was also carried out for approximation of allele sizes and on a zero DNA control. PCR products were electrophoresed through a 4.75% urea polyacrylamide gel and analysed with GeneScan 3.1 software. GS2500 ROX standard is depicted red, PCR products are blue.

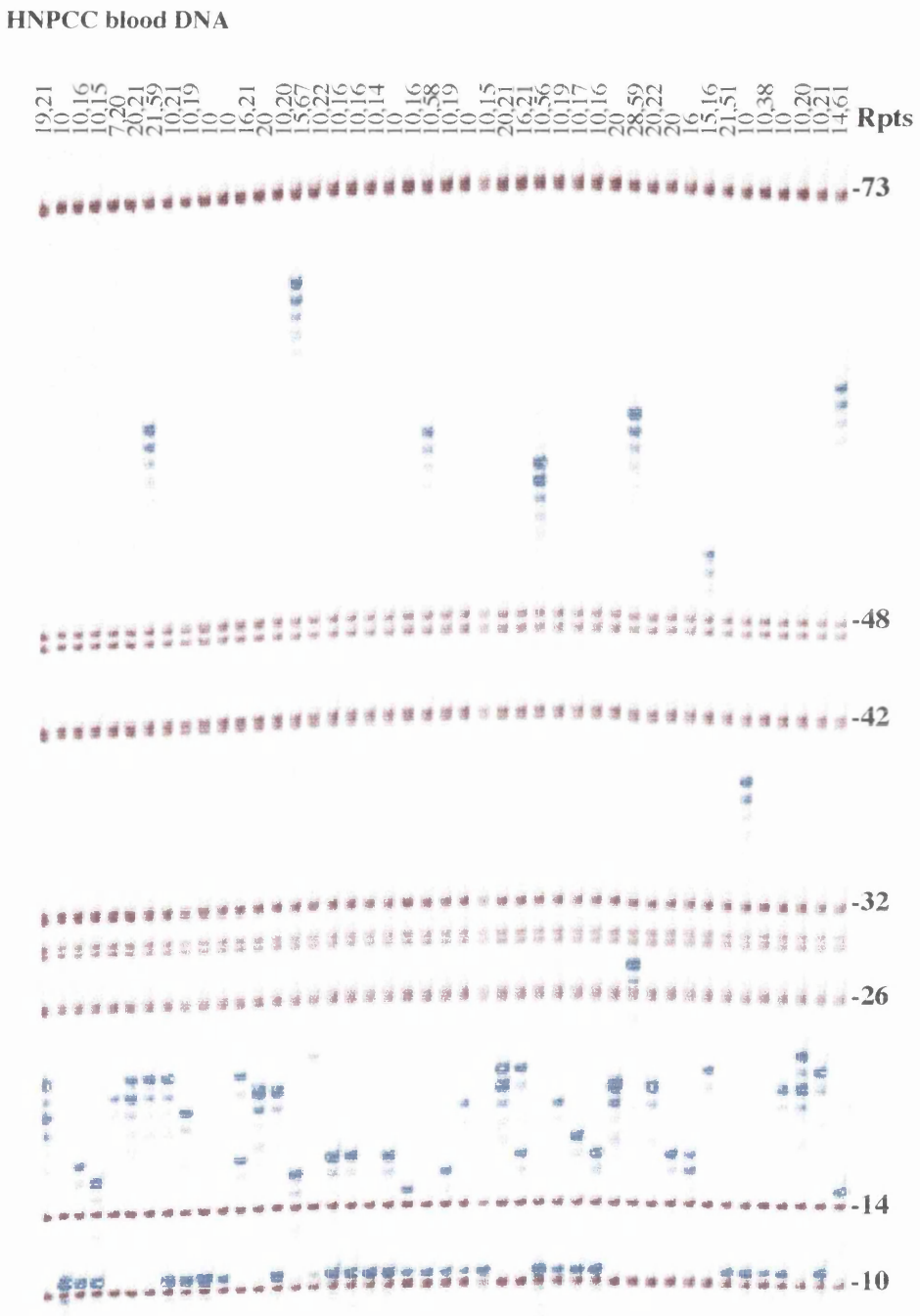

Figure 3.4 Fluorescent allele typing at the *ERDA1* locus. Each typing reaction was carried out with primers ERDA-A-F/DR. Control PCR on genomic DNA DAMO (19,62) and GOMC (22,60) was also carried out for approximation of allele sizes and on a zero DNA control. PCR products were electrophoresed through a 4.75% urea polyacrylamide gel and analysed with GeneScan 3.1 software. GS2500 ROX standard is depicted red, PCR products are blue.

HNPCC blood DNA

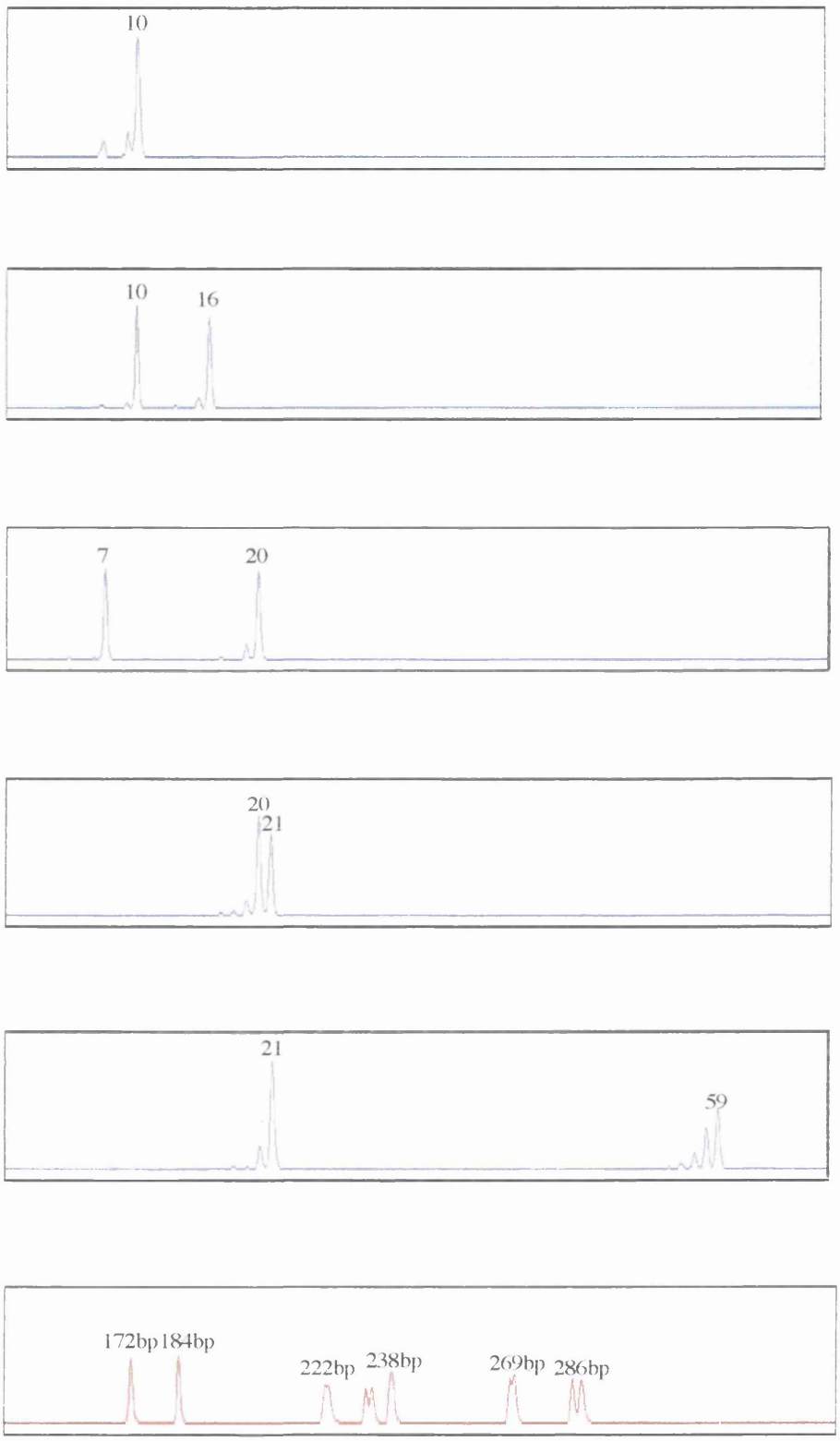

Figure 3.5 GeneScan profiles of *ERDA1* fluorescent PCR products. GeneScan analysis of fluorescent PCR products electrophoresed on an automated sequencing machine produced separate profiles for each sample. The PCR product is depicted blue and the GS2500ROX standard depicted red. The fluorescence is detected as a peak and the size of the product recorded in bp which is converted to repeats. Sizes of the alleles in repeats is depicted.

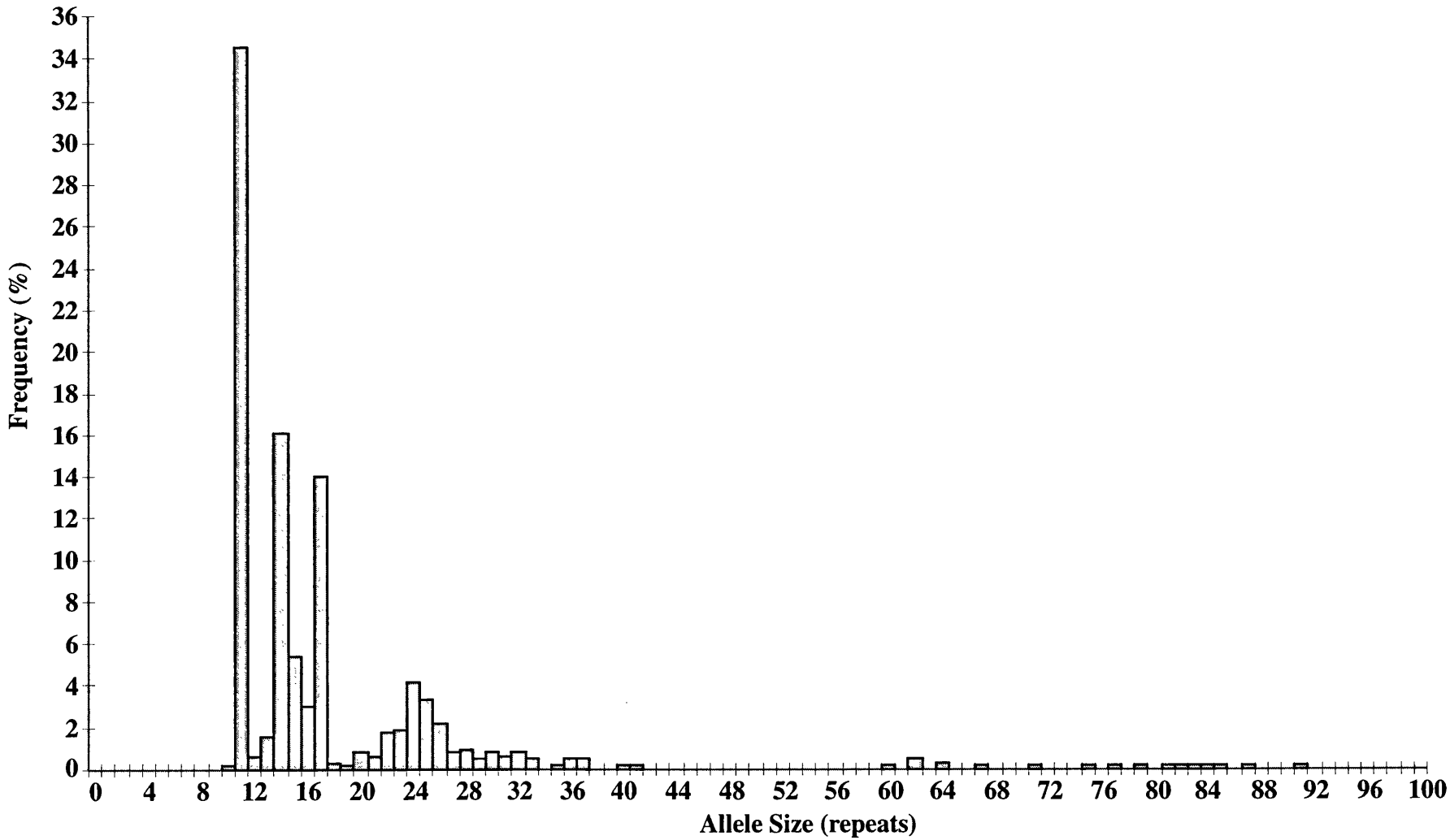

Figure 3.6 Distribution of allele frequencies at the *CTG18.1* locus in male individuals. 628 chromosomes were typed at the *CTG18.1* locus using primers SEF2-C-F/BR. The most common allele is 11 repeats with a frequency of 34.6%. Frequency of expanded alleles (>50 repeats) is 2.9%.

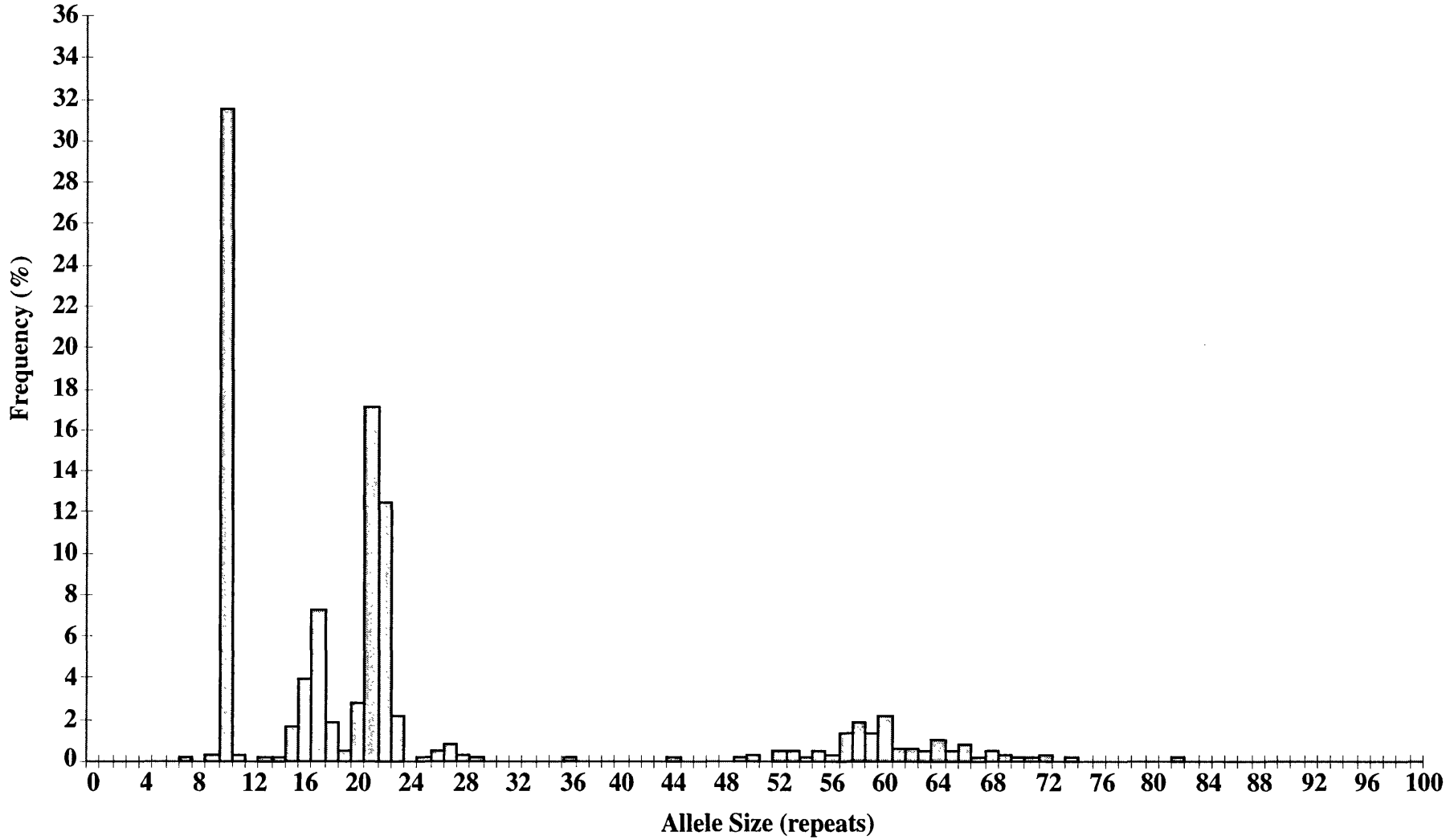

Figure 3.7 Distribution of allele frequencies at the *ERDA1* locus in male individuals. 616 chromosomes were typed at the *ERDA1* locus using primers ERDA-A-F/DR. The most common allele is 10 repeats with a frequency of 31.5%. Frequency of expanded alleles (>50 repeats) is 15.1%.

Chapter 4

Male germline repeat behaviour at the *CTG18.1* and *ERDA1* loci

4.1 Introduction

Typing of individuals from affected families of trinucleotide repeat diseases explained the phenomenon of anticipation occurring in transmission. Adaptation of the SP-PCR technique to microsatellite loci, predominantly the *DM-1* locus, then allowed the precise and differing behaviour of these repeats to be identified (Monckton *et al.*, 1995, Martorell *et al.*, 1998, Monckton *et al.*, 1999). Huge alterations in repeat size with a bias toward expansions could be observed at the *DM-1* locus in the sperm of even asymptomatic individuals, whereas the blood DNA from the same individual is relatively stable (Monckton *et al.*, 1995). Muscle also showed large CTG expansions at the *DM-1* locus that correlated with the disease phenotype (Monckton *et al.*, 1995).

Recent comparisons of the relative stability of the various CAG/CTG trinucleotide repeats in transmission suggests that flanking sequence is a major contributor (Brock *et al.*, 1999). A positive correlation was found between increasing %GC content in the flanking sequence and expandability. The limited data available for *ERDA1* suggests it has the lowest expandability in male transmission, whereas *HD, SCA2, SCA7* and *DM-1* have high expandabilities correlating with high %GC content. For this reason, we sought to further define the behaviour at the *ERDA1* locus and generate data for the *CTG18.1* locus in the male germline and to compare this with data from other trinucleotide repeat loci utilising the sensitive SP-PCR technique.

4.2 Results

4.2.1 SP-PCR at the *CTG18.1* and *ERDA1* loci

Following typing of male semen donors at the *CTG18.1* and *ERDA1* loci (Section 2.2), candidates were selected to provide a broad range of data comprising normal and expanded alleles. Good quality sperm DNA was isolated, restriction digested and a preliminary SP-PCR was carried out at either *CTG18.1* or *ERDA1* (Section 2.3) on a range of concentrations in order to optimise the SP-PCR reactions for approximately 1 and 100 molecules per

reaction (Figure 4.1). The purpose was to observe any frequent small length changes (<5 repeats) at the single molecule level and rarer, large length changes (>5 repeats) occurring at the level of 100 molecules per SP-PCR reaction. Reactions were scored positive according to whether a lane displayed one or more bands at the low concentration, separately for each allele. From the preliminary SP-PCR, 8 reactions of the low concentration were scored and the frequency of negative reactions (p(0)) was calculated as :

total-positive/total

The number of molecules observed at this concentration of DNA per SP-PCR reaction, was calculated by applying p(0) to the equation derived from the Poisson distribution:

no. molecules per reaction = -ln p(0)

This value indicates the approximate number of molecules on average per reaction at this concentration of DNA for a given allele. Therefore, the concentration of DNA needed for 1 molecule and 100 molecules per reaction can be calculated and an increased number of reactions carried out to observe any rare, large length changes. From these blots, a more accurate value was calculated for the number of molecules analysed and the mutation rate and average length change was then estimated. It was also important to monitor contamination and so 32 negative DNA SP-PCR reactions were carried out with the same reagents as the sperm DNA samples. Each set of reactions was only included for analysis if no contamination was detected.

Blood DNA was not available to estimate the progenitor allele sizes of the male donors, so DNA from the somatic cell fraction of the sperm lysis protocol was isolated (Section 2.1.2) and fluorescent PCR carried out in order to type these individuals (Sections 2.2.2 and 2.2.4). A total of 15 semen samples underwent SP-PCR analysis at the *CTG18.1* locus (Table 4.1) and 8 at the *ERDA1* locus (Table 4.2). An average of 7,890 sperm were analysed per allele, per individual for large concentration SP-PCR. Examples of large scale *CTG18.1* SP-PCR with low concentration DNA (Figure 4.2) and high concentration DNA are displayed (Figure 4.3). Normal alleles less than 50 repeats at the *CTG18.1* locus have low but detectable large length change mutation rates of 0-0.02%. The two mutations that were detected were expansions of 11 and 15 repeats from 35 and 38 repeat alleles respectively. A moderately expanded allele of 60 repeats had two expansions of 6 and 9 repeats, and thus a mutation rate of 0.02%. Alleles of 70-80 repeats had mutation rates of 0.22-0.73% for large length changes and 80-91 repeat alleles, 0.04-1.57%. The alleles in the 80-91 repeat range are variable as an 84 repeat had a very low mutation rate of 0.04% and an 87 repeat (GS12) had a much higher

mutation rate than a 91 repeat (1.57% and 0.52% respectively). The average length changes for these large mutations suggest that the normal and moderately expanded alleles increase by 0.002 repeats per sperm and the large alleles have an increasing number of contractions and expansions with increasing repeat length resulting in average length changes ranging from –0.052 to +0.592 (87 repeat of GS12). Therefore the combination of increased mutation rate and average length change of the 87 repeat would suggest that some other *cis* or *trans*-acting factor could be contributing to its increased instability.

Figure 4.4 depicts low and high concentration SP-PCR at the *ERDA1* locus in three different samples with increasing upper allele sizes. The *ERDA1* locus is much more stable than the *CTG18.1* locus and large length changes are rare. Normal alleles less than 50 repeats are very stable, although one expansion of 25 repeats was observed from a 29 repeat allele (mutation rate 0.01% (Table 4.2)). Only three other large length changes were observed in the remaining samples, expansions of 23 and 34 repeats from the 59 repeat allele and 28 repeats from the 67 repeat allele (mutation rates of 0.02% for both alleles). No conclusions can be drawn from the average length changes due to the small number of samples that showed mutations. Sample GS12 is the only sperm sample to be analysed at both loci. Its 87 repeat at the *CTG18.1* locus has a higher mutation rate and increased average length change for large length mutations, and its 67 repeat at the *ERDA1* locus is one of only two expanded alleles that demonstrate mutations. Therefore it seems possible that there is some other factor that could be contributing to increased instability at the *CTG18.1* and *ERDA1* loci in this sample.

4.2.2 Single molecule fluorescent PCR at the *CTG18.1* and *ERDA1* loci

SP-PCR analysis provides a generalised picture of the behaviour of repeats but small length changes are difficult to measure accurately. Therefore by utilising GeneScan analysis which allows accurate sizing of differences between alleles in the same sample, mutation rates and average length changes can be calculated exactly. First round SP-PCR reactions of low concentration DNA were diluted 1:100 and cycled through a second round PCR reaction incorporating a fluorescent forward primer. PCR products were run on a 4.75% polyacrylamide gel with GS2500 ROX standard and analysed with GeneScan 3.1 software. On average, 117 molecules per allele were analysed, the peak with the greatest area taken to indicate allele size.

Figures 4.5 and 4.6 depict examples of GeneScan gels with second round single molecule PCR products at either the *CTG18.1* or *ERDA1* loci. PCR products are blue and size standard red. The lower alleles are stable, displayed as a clear sharp band. The upper alleles however, have broader profiles made up of multiple bands that increase with increasing allele

length. These extra bands that accompany the main band are PCR artefacts called "stutter" and mainly precede, but sometimes are larger in size than the main PCR product. Analysis of the profile of each lane (and hence reaction) means negative PCR reactions can be clearly scored and the presence of greater than one allele can be easily detected (Figure 4.7). As depicted in Figure 4.7, a single molecule will produce a peak with a series of progressively smaller peaks preceding it. However, if any peak preceding the largest peak has a larger height than expected of stutter, this is also classed as a real peak. Figures 4.5 demonstrates how increasing allele size leads to increasing mutation rate at the *CTG18.1* locus. The majority of allele length changes are within 1 repeat in sample GS153 (11, 38) but an expansion of 6 repeats can be seen in the final lane. Sample GS166 (11, 71) however has a larger upper allele and hence changes of 1-2 repeats are more common. The frequencies of these alleles are represented in Figure 4.8 which shows how increasing repeat length leads to an increasing spread of alleles within the population of sperm. Figure 4.6 shows the same behaviour at the *ERDA1* locus. Sample GS174 has allele sizes of (16, 29) and no changes can be observed in 106 molecules analysed. Sample GS110 has larger alleles of (10, 74) and the upper allele can be seen to be moderately unstable with frequent changes of 1-2 repeats, and deletions of 4 and 5 repeats are evident.

By measuring the frequency of mutations at these two loci in various semen samples, mutation rates and average length change were calculated. Firstly the number of molecules analysed was calculated for each allele in the sample based on the observations on the GeneScan gel. The total number of observed positives for each allele was divided by the total number of observed positives analysed, then multiplied by its difference from the progenitor allele (+/-) to give an average length change, in repeats. Displayed in Tables 4.1 and 4.2 are the number of molecules per reaction, total number of molecules analysed, mutation rate and average length change for both alleles of each sample analysed at the single molecule level.

The mutation rates and average length changes at both loci are displayed in Figures 4.9, 4.10, 4.11, and 4.12. Figure 4.9 is a scatter graph depicting the mutation rates of alleles at the *CTG18.1* locus from 11 to 91 repeats in length. There is clearly an increasing mutation rate with increasing allele length, displayed as a sigmoidal curve. Alleles of 11-17 repeats are very stable with no mutations, alleles less than 36 repeats have increased mutation rates of between 3 and 24 % and alleles from 67-91 repeats have elevated mutation rates of between 75 and 89%. 95% confidence intervals for the mutation rate, taking into account 95% confidence intervals for the number of sperm analysed, are also plotted for each allele where possible. Plotting the average length change at the *CTG18.1* locus demonstrates that as the alleles increase in length, they do not necessarily expand (Figure 4.10). In fact, the majority of alleles are contracting. Alleles less than 40 repeats, which can have mutation rates of up to

24%, seem to have relatively equal numbers of expansions and contractions and hence have only moderate overall contraction rates. Alleles from 67-91 repeats on the whole tend to contract by up to two repeats on average. The alleles from 60-71 repeats have low average length changes, either positive or negative, which could help to explain why 40-60 repeats are so rare in the samples typed. There is a greater probability that mutations would be less than 5 repeats and so new alleles would remain within the same range. The 87 repeat has an unusually large positive average length change compared to the other alleles in this repeat range, confirming the large length data from the same sample and increasing the positive average length change (Table 4.1).

The mutation rates of *ERDA1* alleles are displayed in Figure 4.11. Although there are a decreased number of alleles analysed at this locus, there is a clear difference from the *CTG18.1* locus. The alleles are more stable, with alleles of equivalent length possessing lower mutation rates. Whereas at the *CTG18.1* locus alleles of 20-30 repeats are mutating, at the *ERDA1* locus alleles are stable until the 30-40 repeat range. Again there is a lack of alleles available between 40-50 repeats, but alleles of 81 repeats have a mutation rate as low as 59%. At the *CTG18.1* locus, an equivalently sized allele would have a mutation rate of approximately 80%. The average length change at this locus is plotted in Figure 4.12. The mutation rate demonstrates that this locus is particularly stable and so any large changes would be unexpected. The greatest average length change was in fact from the largest allele of 81 repeats, but still only moderate at -0.6 repeats per sperm. All alleles that did mutate did so with a bias toward contractions at this locus.

To address the question of meiotic drive at the *CTG18.1* and *ERDA1* loci, the single molecule observations of each sample underwent a Chi-squared test and no significant difference was found between the number of positives for each allele. i.e. there are equal numbers of the small versus the large allele in each sample.

4.3 Discussion

Direct comparison of the two loci, *CTG18.1* and *ERDA1*, with each other and the *DM-1* locus, demonstrates that the same sequence of repeat can behave very differently. Single molecule data from the *DM-1* locus which also has an expandable CTG repeat (Appendix 2) shows a particularly unstable locus heavily biased toward expansion in the male germline. Mutation frequencies become measurable in the mid 20 repeat range with a mixture of expansions and contractions until the alleles reach approximately 30 repeats in length. Then the alleles start to expand much more frequently until they reach approximately 35 repeats whereby they achieve

mutation rates of 60%. SP-PCR analysis at this locus demonstrates that alleles greater than 50 repeats (defined as expansion alleles) have essentially a 100% mutation rate, 99% biased toward expansions (Monckton *et al.*, 1995).

CTG18.1 and *ERDA1* however seem to be more stable than the *DM-1* locus, but still possess underlying mutation rates of detectable levels. SP-PCR analysis demonstrates that large length changes are rare at *CTG18.1* even with large alleles of up to 91 repeats and even less evident at the *ERDA1* locus. Again, at the single molecule level, GeneScan analysis confirms that *CTG18.1* has a lower mutation rate than *DM-1* and a higher mutation rate than *ERDA1*. In fact, all three loci demonstrate the pattern of increasing mutation rates with increasing allele length, but the nature of these mutations is what makes these loci distinct. *DM-1* has a heavy bias toward expansions, 30-40 repeat alleles mutating at up to 60% frequency, leading to the disease alleles (>50 repeats) giving rise to the DM phenotype. The *DM-1* CTG repeat is located in the 3' UTR of the *DM-1*PK gene and the putative promoter of the *Six5* gene. The flanking sequence of this repeat is high in %GC content and also within the CpG island for *Six5*. The *CTG18.1* CTG repeat is located in the third intron of the *SEF2-1* gene which has a lower flanking GC content and a lack of extensive human genomic sequence for the *SEF2-1* gene means that the location, if any, of a CpG island is yet to be determined and related to the instability of the CTG repeat. The *ERDA1* locus has been sequenced in a BAC clone and NIX analysis (Williams *et al.*, 1998) suggests it is not located within a gene. The fact that the mutation rate declines from *DM-1* to *CTG18.1* to *ERDA1*, with an increasing tendency toward contraction respectively is very likely to be linked to the decreasing GC content in the flanking sequence. The result is reflected in the distribution and frequency of alleles at these loci. The *DM-1* locus has a similar distribution to the *CTG18.1* locus, the most frequent allele being 5 repeats at approximately 35% frequency. The frequency of expanded alleles (>50 repeats) is 1 in 8000. The frequency of expanded *CTG18.1* alleles is 3% and *ERDA1* alleles 15%. The high level of *ERDA1* expanded alleles available is most probably due to their stability. Expanded *ERDA1* alleles have been demonstrated to possess a C polymorphism immediately 3' to the CTG repeat. This polymorphism has been confirmed in the samples analysed by utilising primers differing at the final 5' nucleotide (Section 2.2.5). This allowed selective amplification of alleles possessing the A/C polymorphism and their associated repeat which could be separated by size by agarose gel electrophoresis. All alleles <50 repeats were found to have A 3' to the CTG repeat and all alleles >50 repeats have C 3' to the CTG repeat. Chimpanzee DNA has also been typed for this polymorphism in order to identify the ancestral variant. All the chimpanzee samples indicate small length *ERDA1* alleles associated with the A variant. Therefore it is possible that mutation of this polymorphism from an A to a C could have predisposed a normal allele to a rare expansion, giving rise to the pool of larger alleles >50 repeats.

Data available from 731 transmissions of the *ERDA1* repeat suggests there is a contraction bias of expanded alleles in males and an expansion bias in females, with increasing mutation rate correlating with increasing allele size (Ikeuchi *et al.*, 1998, Deka *et al.*, 1999). The SP-PCR and single molecule data confirms the observations in male transmissions, although the allele length difference is larger in the transmissions than the single sperm data would suggest. Contractions of up to 4 (Ikeuchi *et al.*, 1998) and 15 repeats (Deka *et al.*, 1999) have been observed in male transmissions whereas the largest average length change was of –0.582 repeats (82 repeat allele). However, the single molecule data is effectively measuring the average length change of 6,951 transmissions from this 82 repeat allele (Table 4.1) which provides a more accurate representation of the behaviour of this repeat in the population as a whole. It would be interesting to accurately measure the mutation rates and average length changes in samples with larger alleles that naturally arise in a population such as the Chinese. The higher prevalence of large *ERDA1* alleles in the Chinese could have arisen due to a bottleneck in the evolution of the population, or there may be a predisposition to a higher rate of expansion compared to Europeans.

CTG18.1 alleles are moderately unstable and though expansions are seen, overall there is a tendency toward contraction. Again, there seems to be two sets of alleles less than 100 repeats, but the larger alleles tend to be a little less definite in whether they will expand or contract. These differences could be due to personal genetic differences in the individuals sampled, such as the 87 repeat from sample GS12, which seems to expand more frequently than the 91 repeat for both small and large length changes. Whether these differences are due to polymorphisms in the immediate flanking sequence or variants in other genes is not known, although immediate flanking polymorphisms are currently being investigated at the *CTG18.1* locus. This more varied behaviour is likely to account for the pool of alleles between 60 and 91 repeats and their persistent stable frequency. Large alleles at the *DM-1* locus however are rare, again due to the fact that once these alleles are greater than 50 repeats, they expand rapidly within one generation and are lost through further size increases.

Comparison of the single molecule data for each sample found there was no significant difference in the numbers of small versus large alleles in the sperm of heterozygotes. This would suggest there is no evidence for meiotic drive in the male germline at the *CTG18.1* and *ERDA1* loci.

Comparison of the single molecule data for the *CTG18.1* and *ERDA1* loci with other trinucleotide repeats suggests that *ERDA1* is indeed very stable and *CTG18.1* is moderately unstable. The *DM-1, HD, DRPLA, SBMA, SCA1* and *MJD/SCA3* loci have higher mutation rates for equivalently sized alleles but the closest pattern of mutations is with *MJD/SCA3*

which also demonstrates an increasing frequency of contractions with increasing allele length. *MJD/SCA3* also has a lower %GC content, correlating with its lower expandability (Brock *et al.*, 1999). The data for the two loci studied confirms that *ERDA1* and *CTG18.1* have the lowest expandability correlating with low %GC content in the flanking sequence of any of the CAG/CTG trinucleotide repeats.

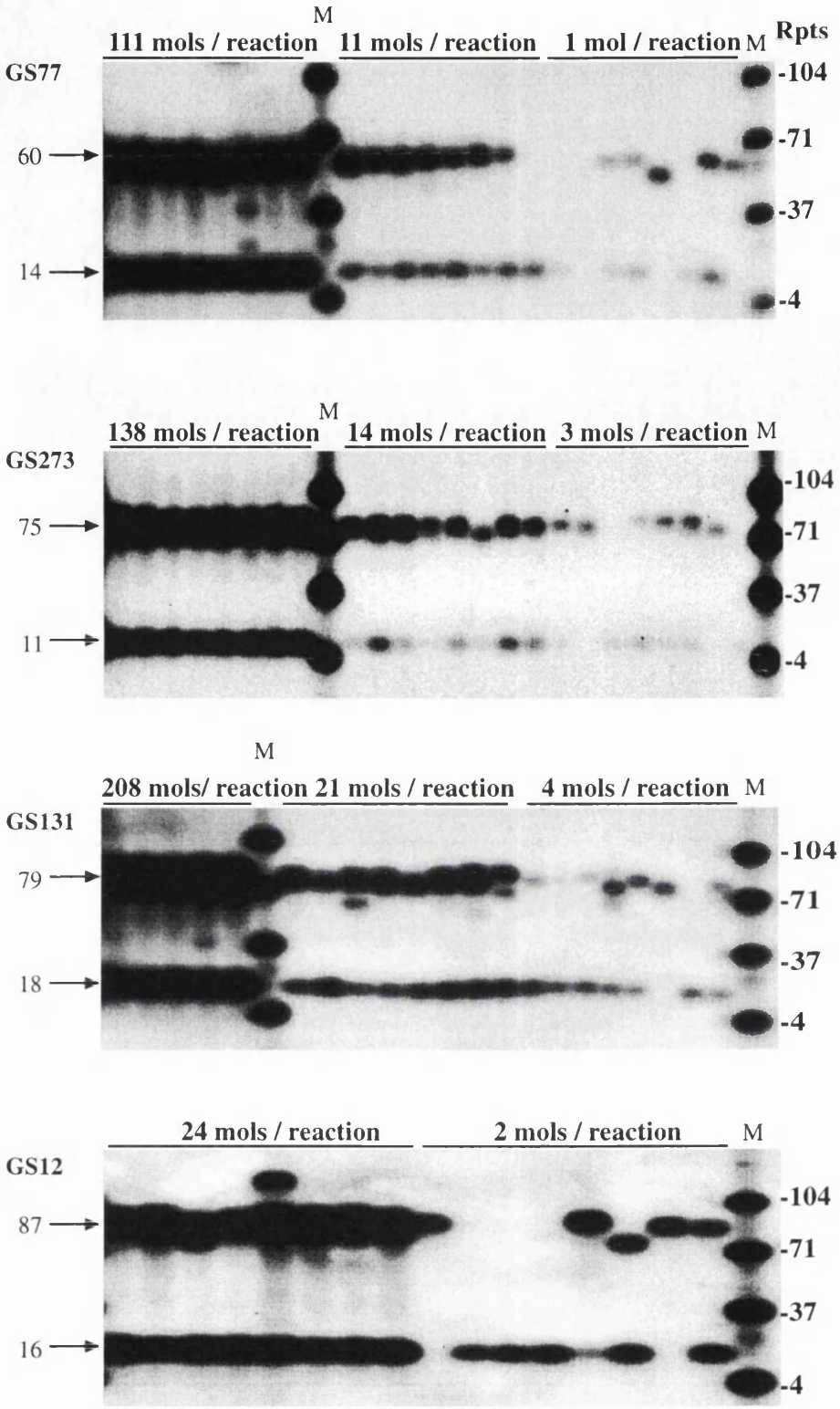

Figure 4.1 Preliminary SP-PCR at the *CTG18.1* locus on sperm. Preliminary SP-PCR at the *CTG18.1* locus on sperm samples GS77 (14,60), GS273 (11,75), GS131 (18,79), GS12 (16,87). Amplisize/2.5Kb marker (M) bands are labelled as equivalent sizes in repeats. Mols = molecules. Positive reactions can be scored and the number of molecules analysed calculated. Frequent small length changes can be observed at the single molecule level and increasing instability with increasing allele size.

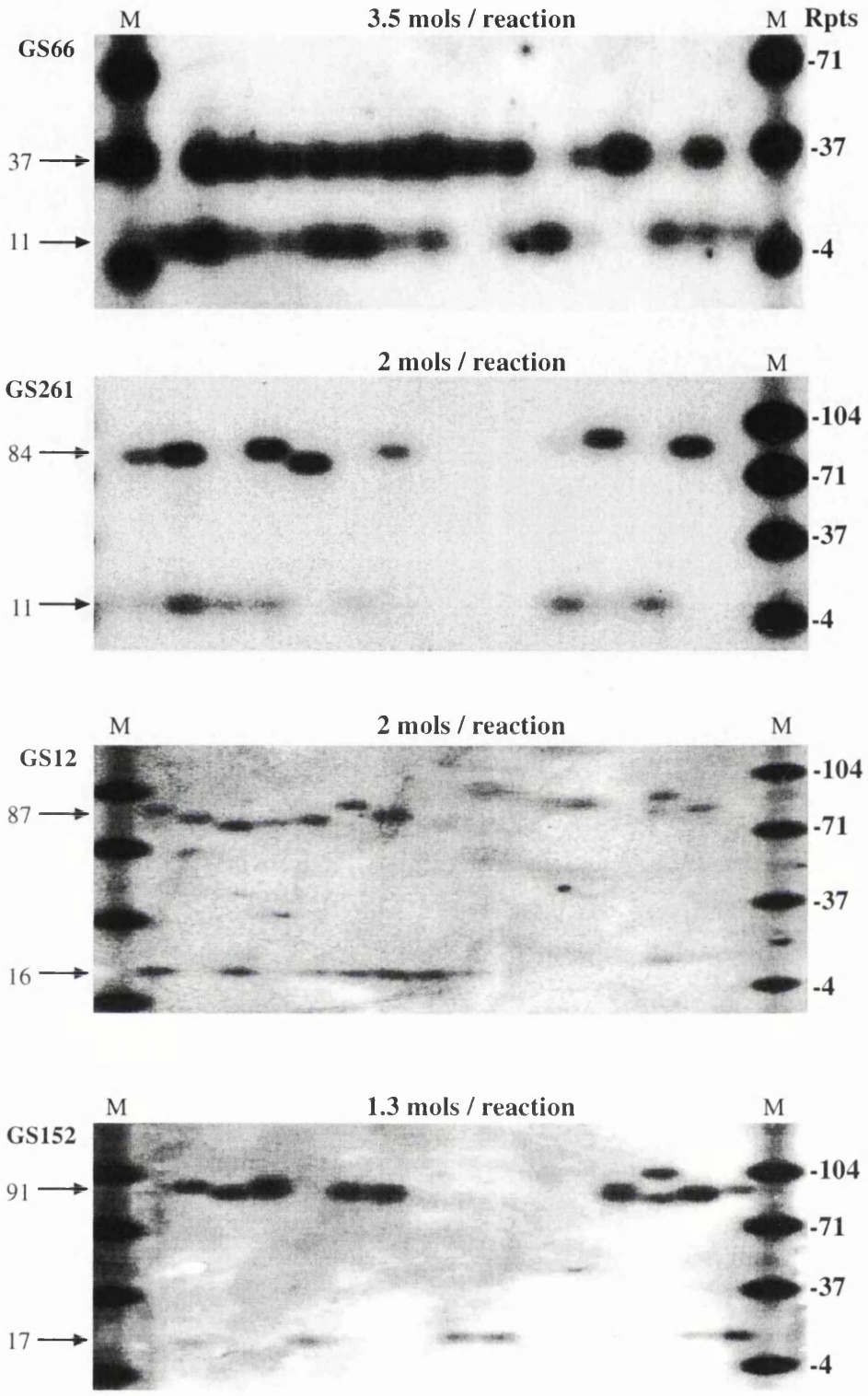

Figure 4.2 Large scale SP-PCR at the *CTG18.1* locus on single molecule DNA isolated from sperm. Single molecule SP-PCR at the *CTG18.1* locus on sperm samples GS66 (11,37), GS261 (11,84), GS12 (16,87) and GS152 (17,91). Amplisize/2.5Kb marker (M) bands are labelled as equivalent sizes in repeats. Mols = molecules. Frequent small length changes can be observed with increasing allele size, although the 91 repeat of GS152 seems more stable than the 87 repeat of GS12.

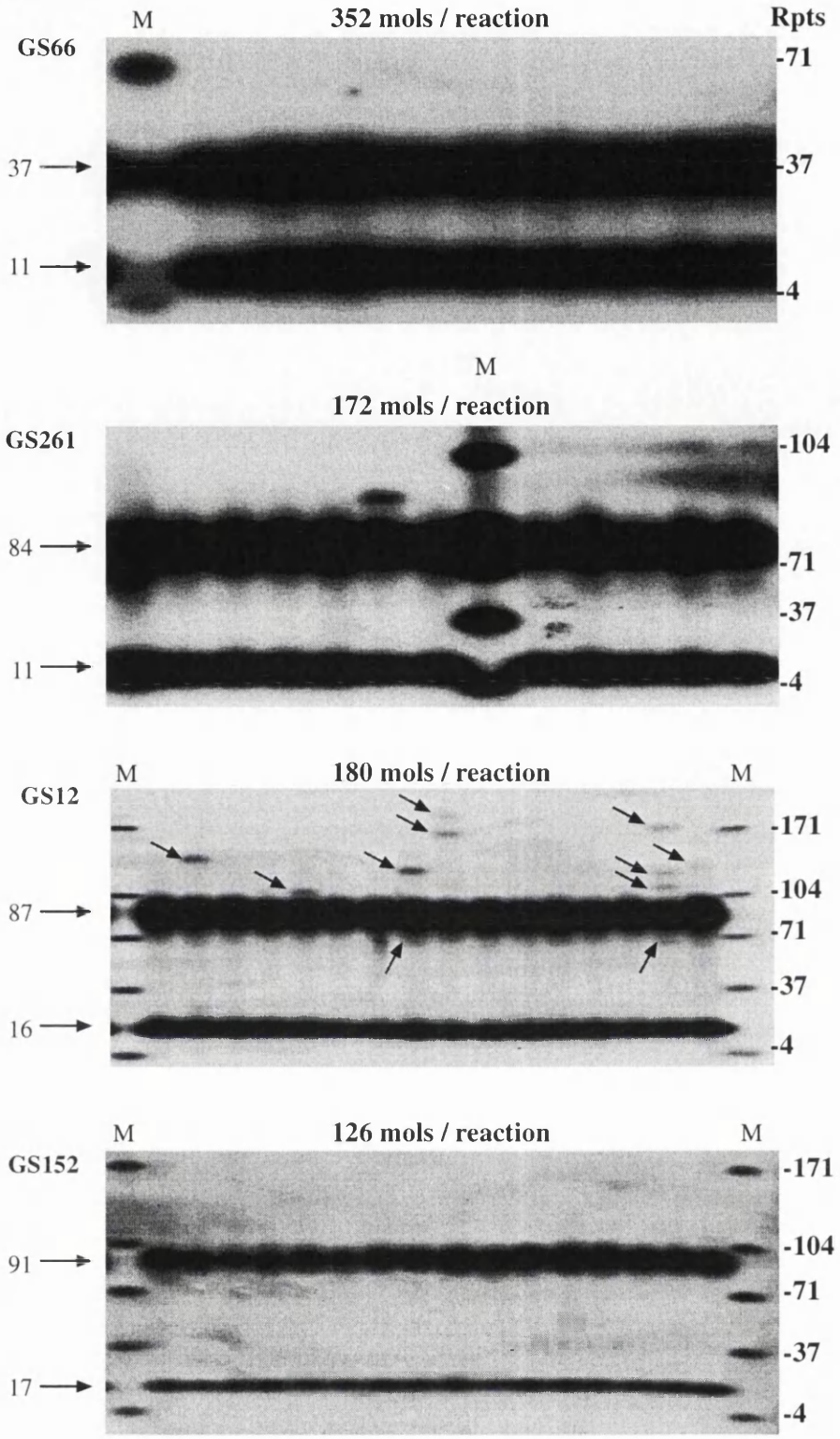

Figure 4.3 Large scale SP-PCR at the *CTG18.1* locus on DNA isolated from sperm. Large scale SP-PCR at the *CTG18.1* locus on high concentration DNA samples GS66 (11,37), GS261 (11,84), GS12 (16,87) and GS152 (17,91). Amplisize/2.5Kb marker (M) bands are labelled as equivalent sizes in repeats. Mols = molecules. Frequent large length changes can be observed from the 87 repeat of sample GS12.

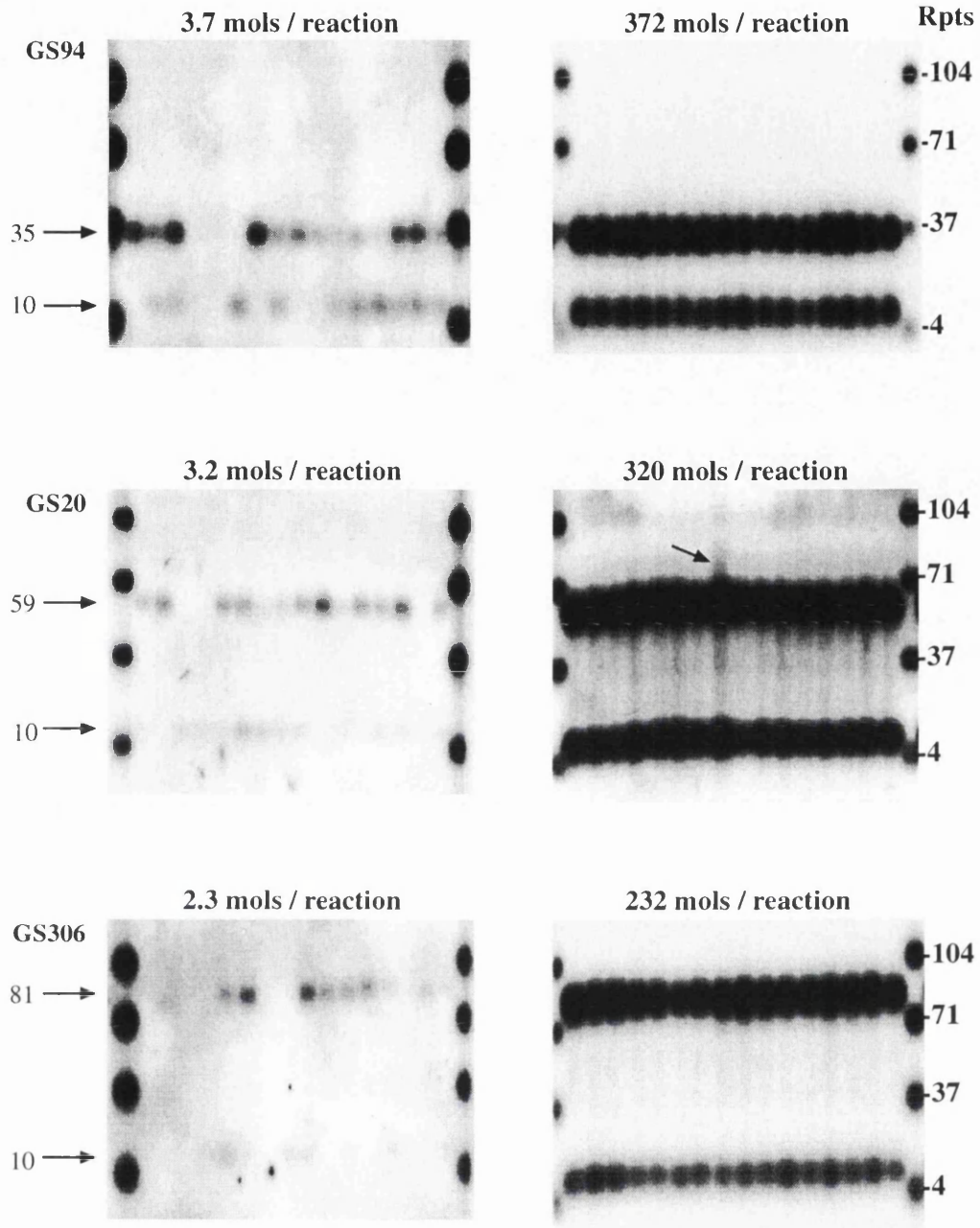

Figure 4.4 Large scale SP-PCR at the *ERDA1* locus on DNA isolated from sperm. Large scale SP-PCR at the *ERDA1* locus on single molecule and high concentration DNA from sperm samples GS94 (10,35), GS20 (10,59) and GS306 (10,81). Amplisize/2.5Kb marker (M) bands are labelled as equivalent sizes in repeats. Mols = molecules. An expansion can be seen in the high concentration DNA of sample GS20 from the 59 repeat allele.

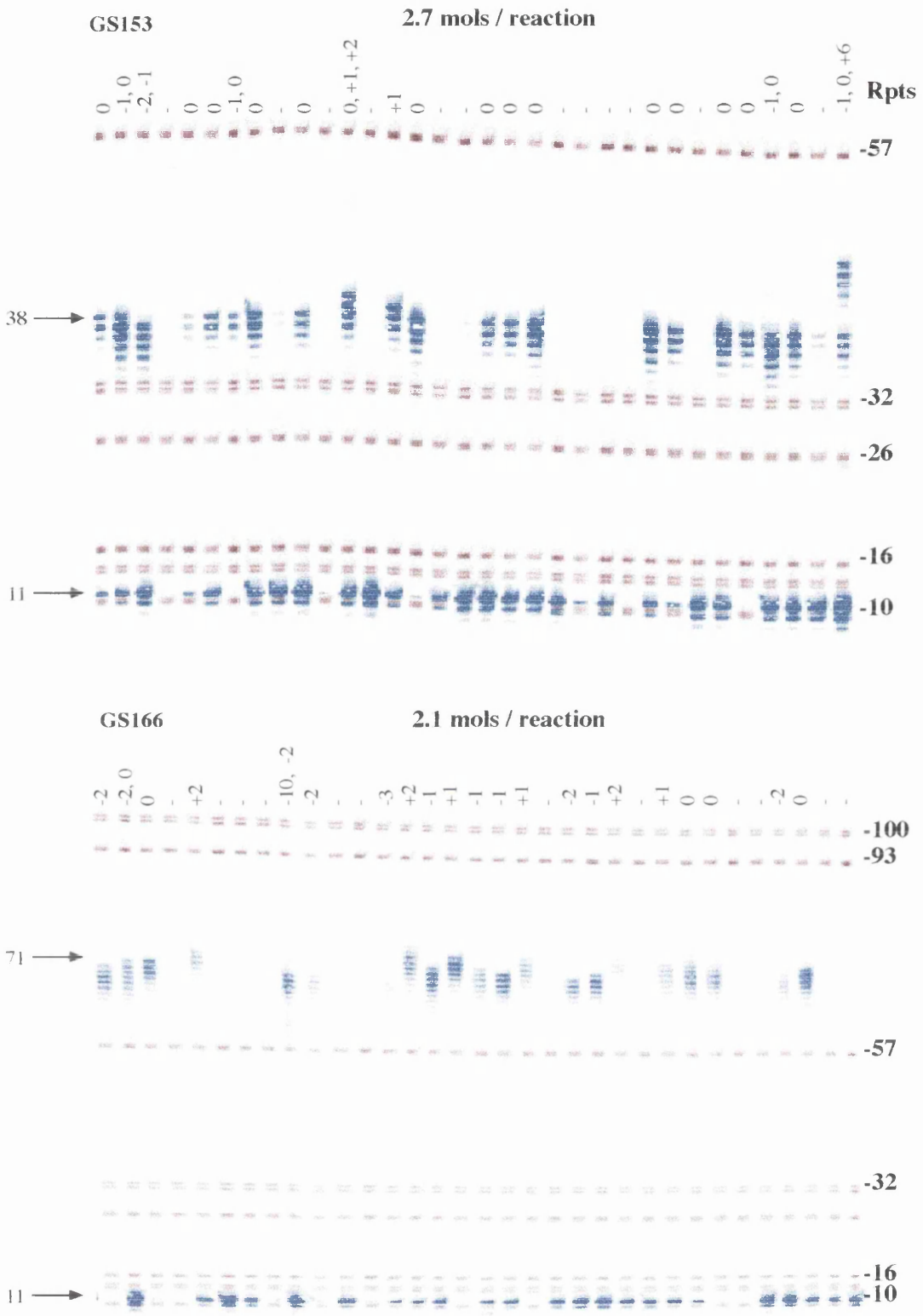

Figure 4.5 Fluorescent PCR on single molecule sperm DNA at the *CTG18.1* locus. Second round PCR was carried out with primers SEF2-C-F/BR on single molecule PCR products from sperm, the products separated on a 4.75% polyacrylamide gel and analysed with GeneScan 3.1 software. GS2500ROX standard is red, PCR products are blue. Mols = molecules. Changes in repeats of the upper allele are depicted above each lane. GS153 (11,38) shows mainly changes of -1 to +1 repeats, although a 6 repeat expansion is evident in the final lane. GS166 (11,71) shows a higher degree of instability with larger length changes of mainly -2 to +2 repeats and a -10 repeat contraction can be detected.

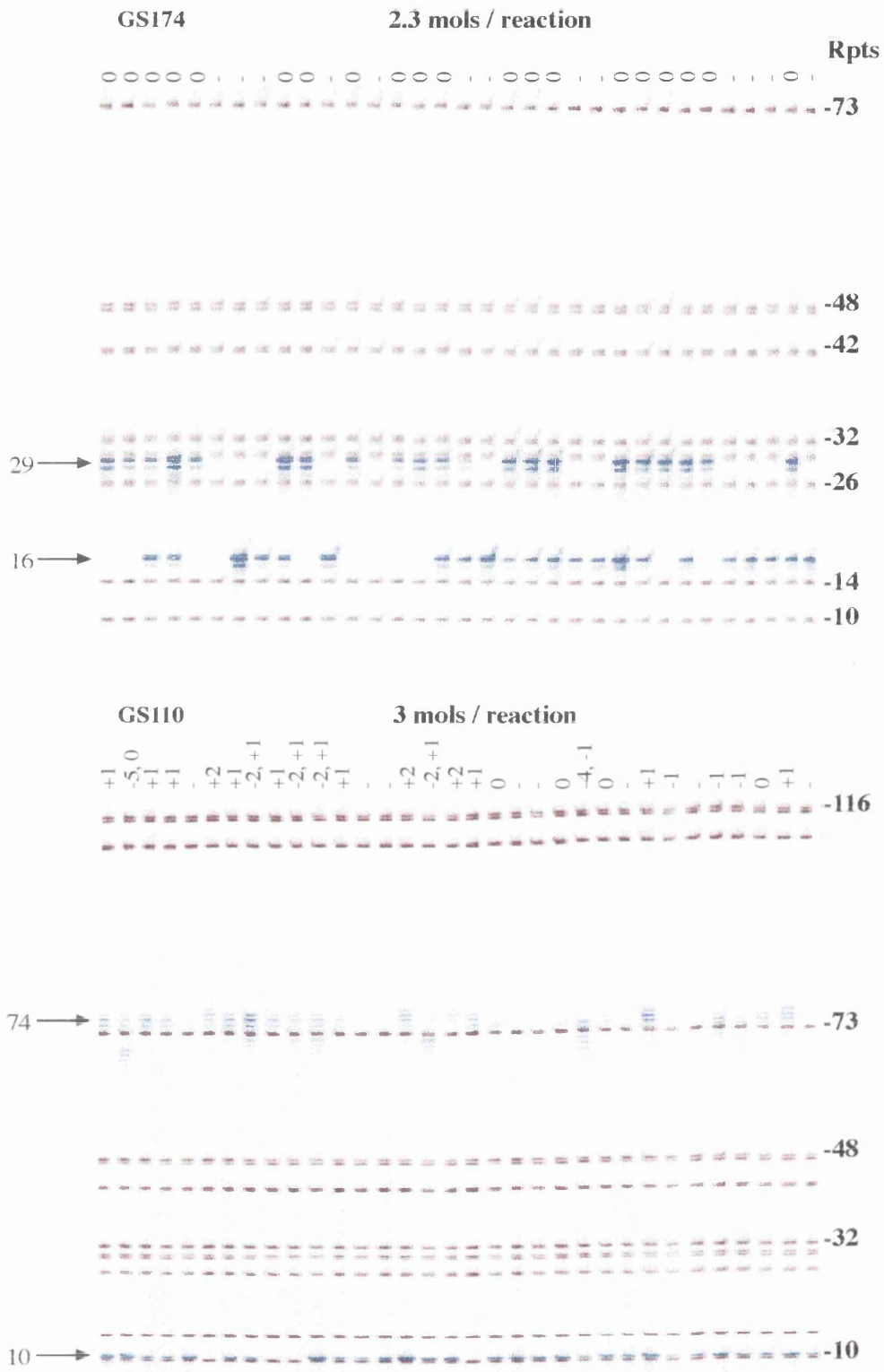

Figure 4.6 Fluorescent PCR on single molecule sperm DNA at the *ERDA1* locus. Second round PCR was carried out with primers ERDA-A-F/DR on single molecule PCR products from sperm, the products separated on a 4.75% polyacrylamide gel and analysed with GeneScan 3.1 software. GS2500ROX standard is red, PCR products are blue. Mols = molecules. Changes in repeats of the upper allele are depicted above each lane. GS174 (16,29)is very stable and shows no mutations. GS110 (10,74) shows a higher degree of instability with length changes of mainly -2 to +2 repeats.

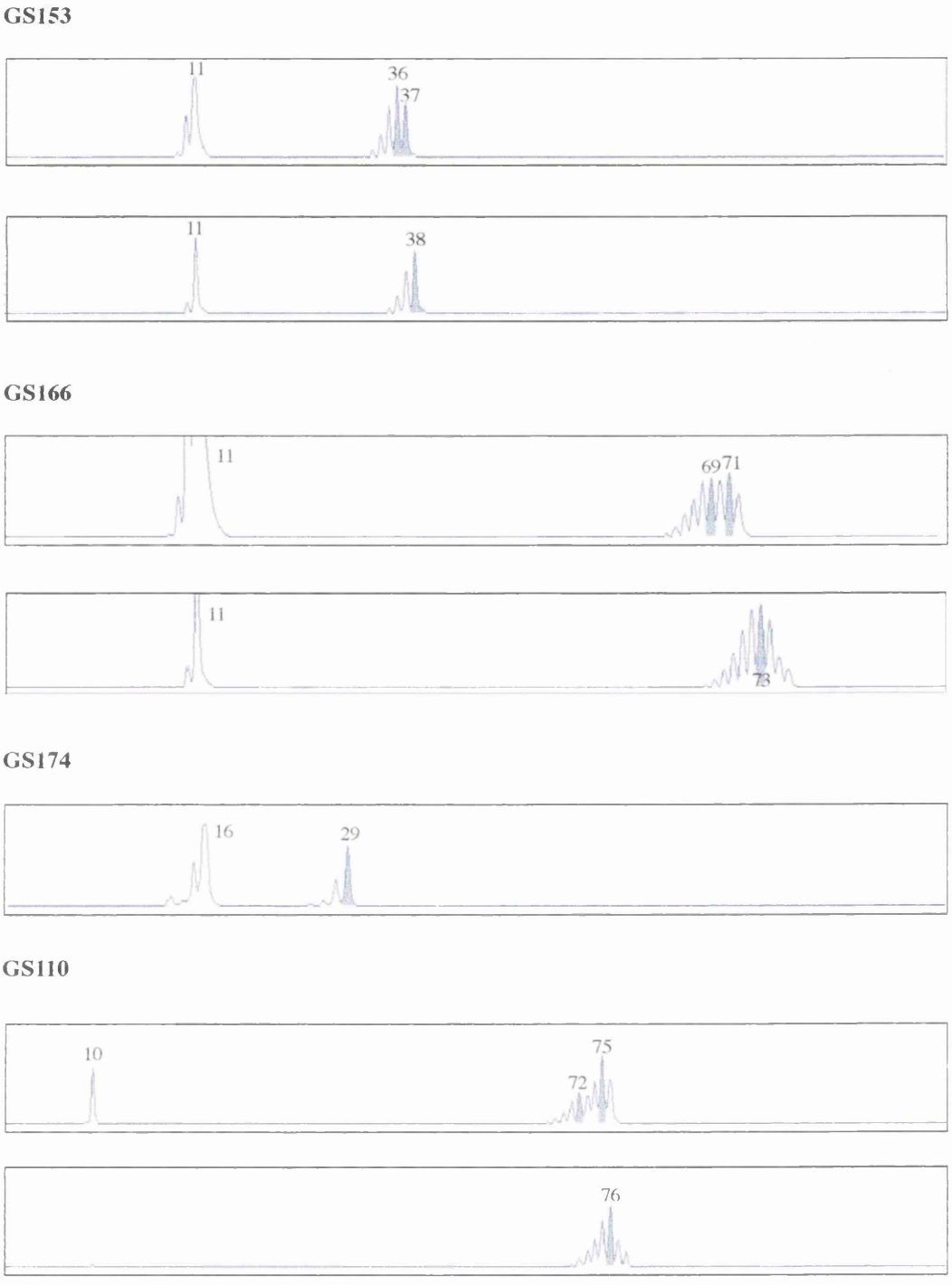

Figure 4.7. GeneScan profiles of second round fluorescent PCR on single molecule PCR products. Depicted are GeneScan profiles of single molecule PCR products from sperm samples GS153 (11,38) and GS166 (11,71) at the *CTG18.1* locus and GS174 (16,29) and GS110 (10,74) at the *ERDA1* locus. Stutter bands can be seen as smaller peaks preceding the main peak and in larger alleles, as a 'hedgehog'. *CTG18.1* alleles have a larger spread of stutter bands than *ERDA1* alleles with increasing repeat length. Filled peaks depict the upper allele for each sample and the respective repeat sizes are labelled.

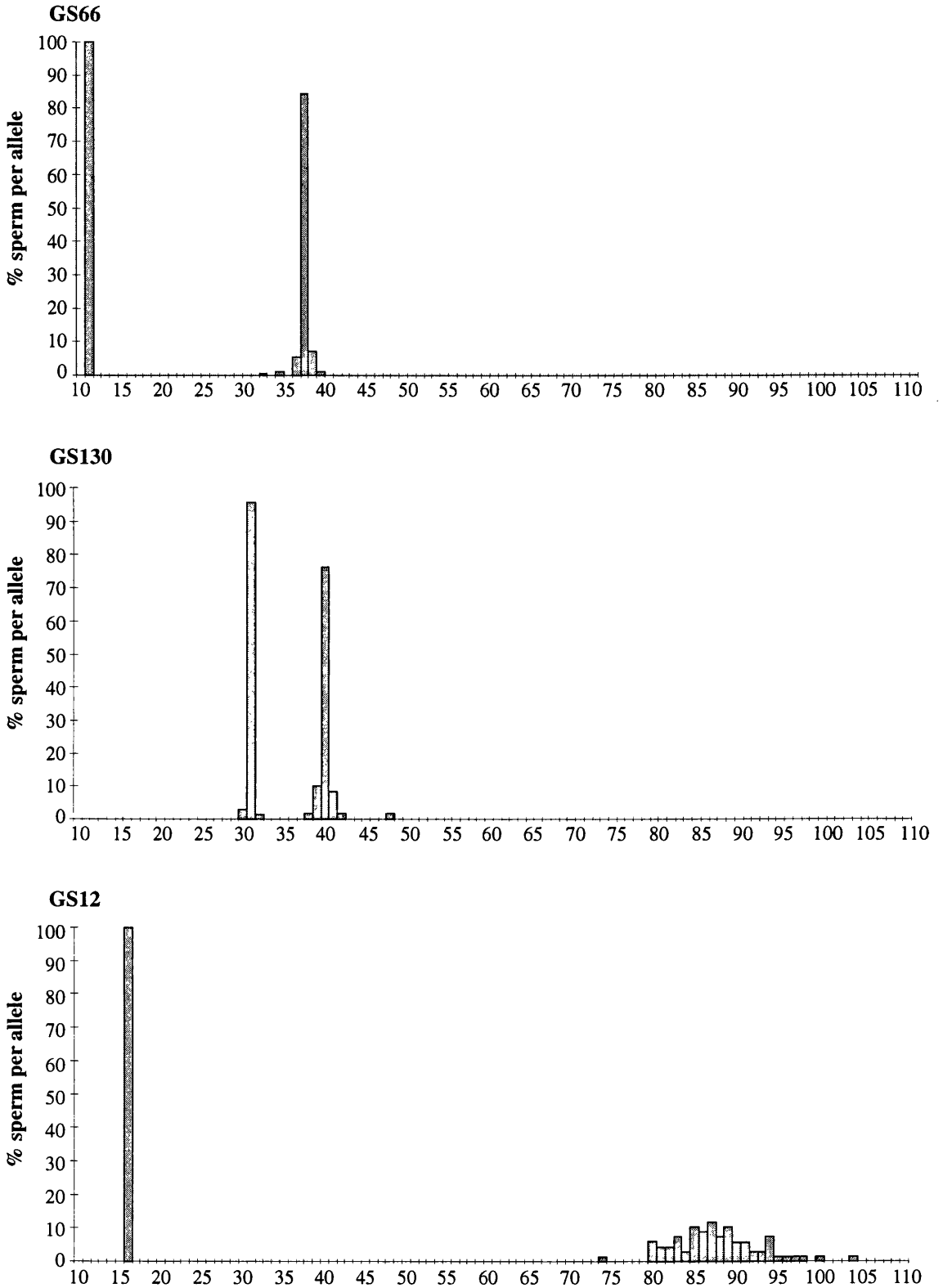

Figure 4.8 Allele frequency distributions from single molecule anlaysis at the *CTG18.1* locus. Displayed are the allele frequency distributions from single molecule PCR analysis of sperm from three heterozygous individuals GS66 (11,37), GS130 (31,40) and GS12 (16,87) at the *CTG18.1* locus. There is increasing mutation rate with increasing allele length resulting in a larger spread of alleles as the repeat increases in length. Between the 30-40 repeat range there is a 20% decrease of the progenitor allele and in sample GS12 the frequency of the progenitor can be seen to be only slightly higher than the mutant alleles surrounding it.

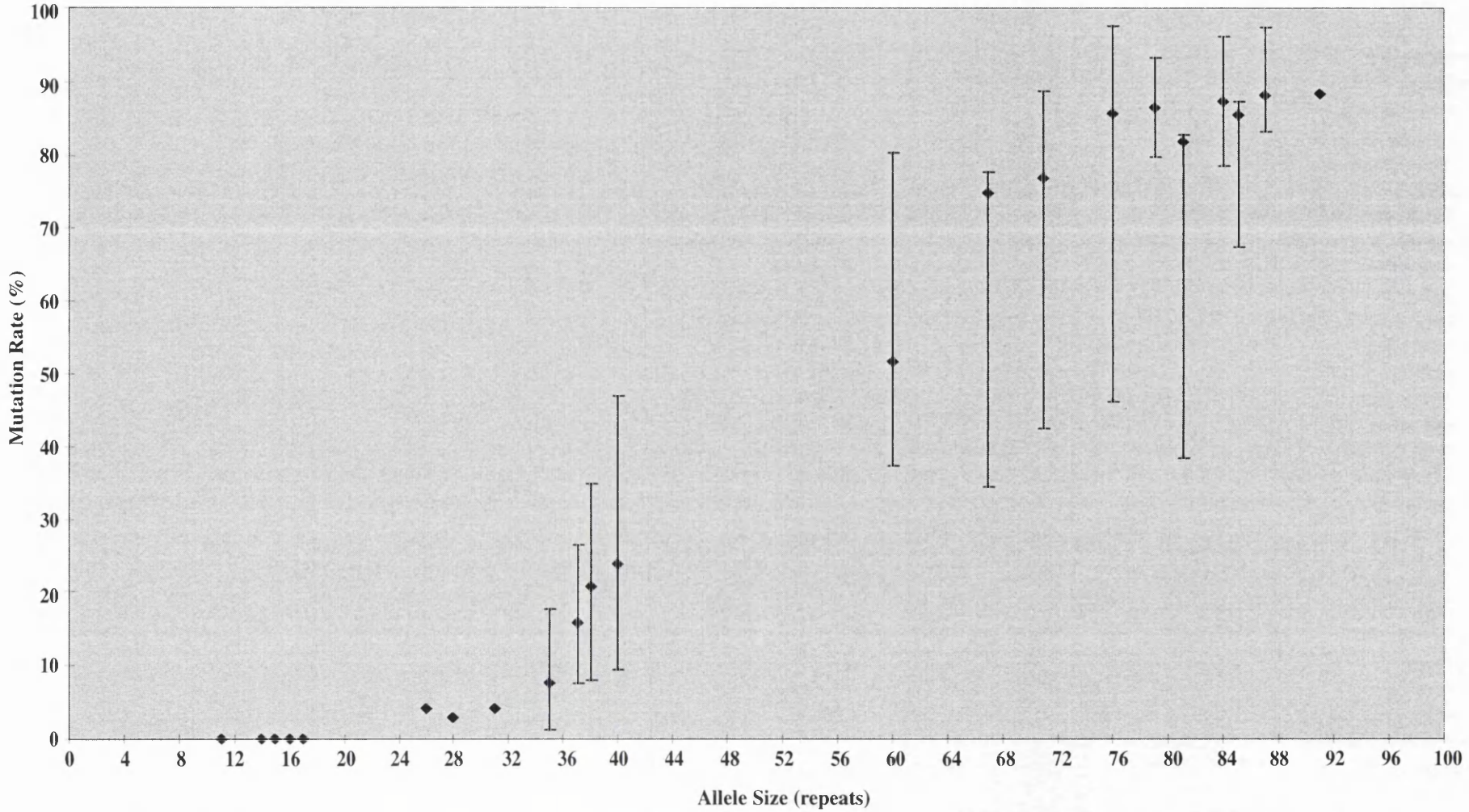

Figure 4.9 Mutation rates of *CTG18.1* alleles in the male germline. Displayed are the mutation rates at the single molecule level of *CTG18.1* alleles in sperm. Where possible, 95% confidence intervals have been calculated and displayed for each allele. The scatter graph shows increasing mutation rate with increasing allele length, depicted as a sigmoidal curve.

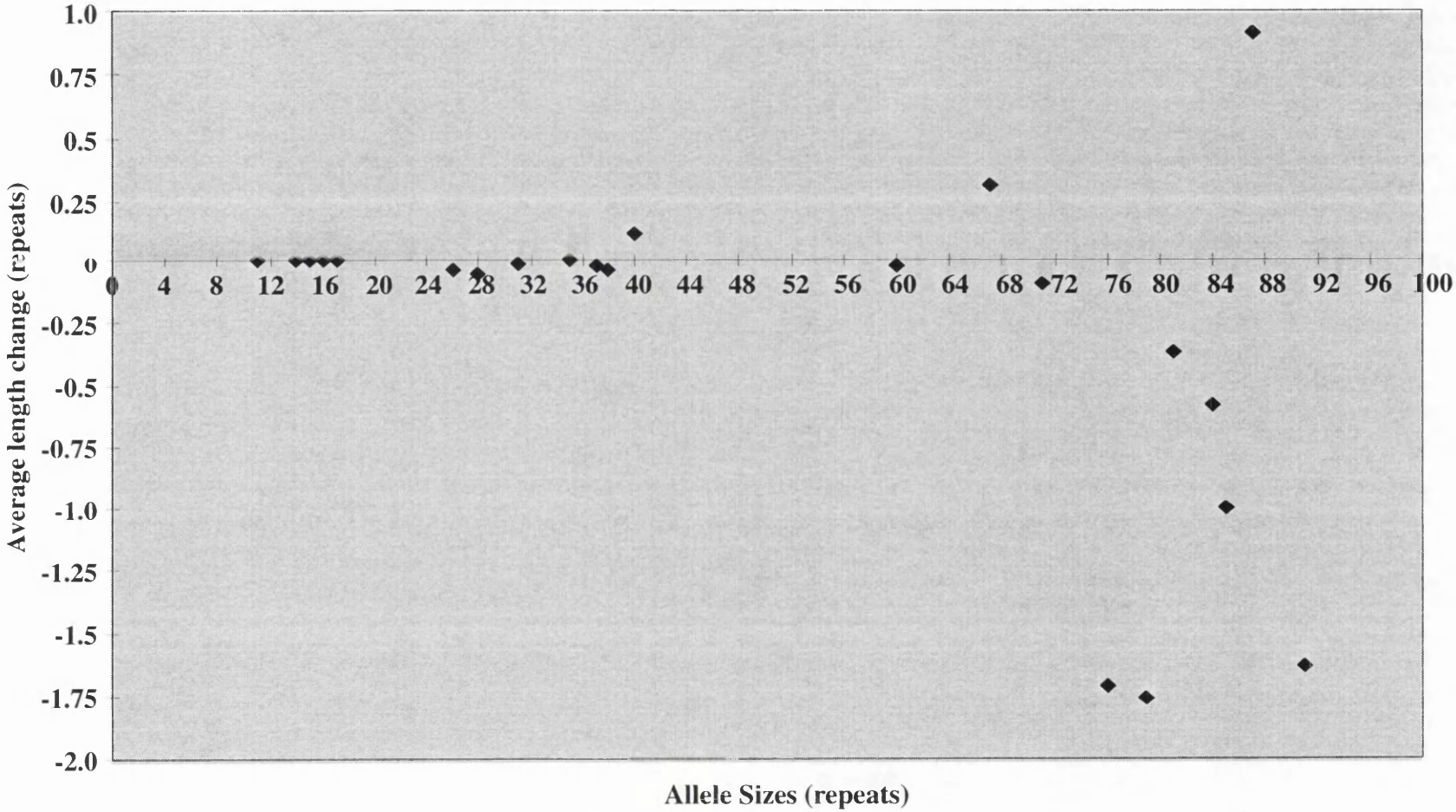

Figure 4.10 The average length change of alleles at the *CTG18.1* locus on single molecule DNA isolated from sperm. Displayed are the average length changes at the single molecule level of *CTG18.1* alleles in sperm. Alleles of 26-40 repeats have an increasing mutation rate but biased toward contractions until the 40 repeat which expands. Between 60 and 91 repeats there seems to be some bias toward contractions, although the 87 repeat is expanding at a high rate.

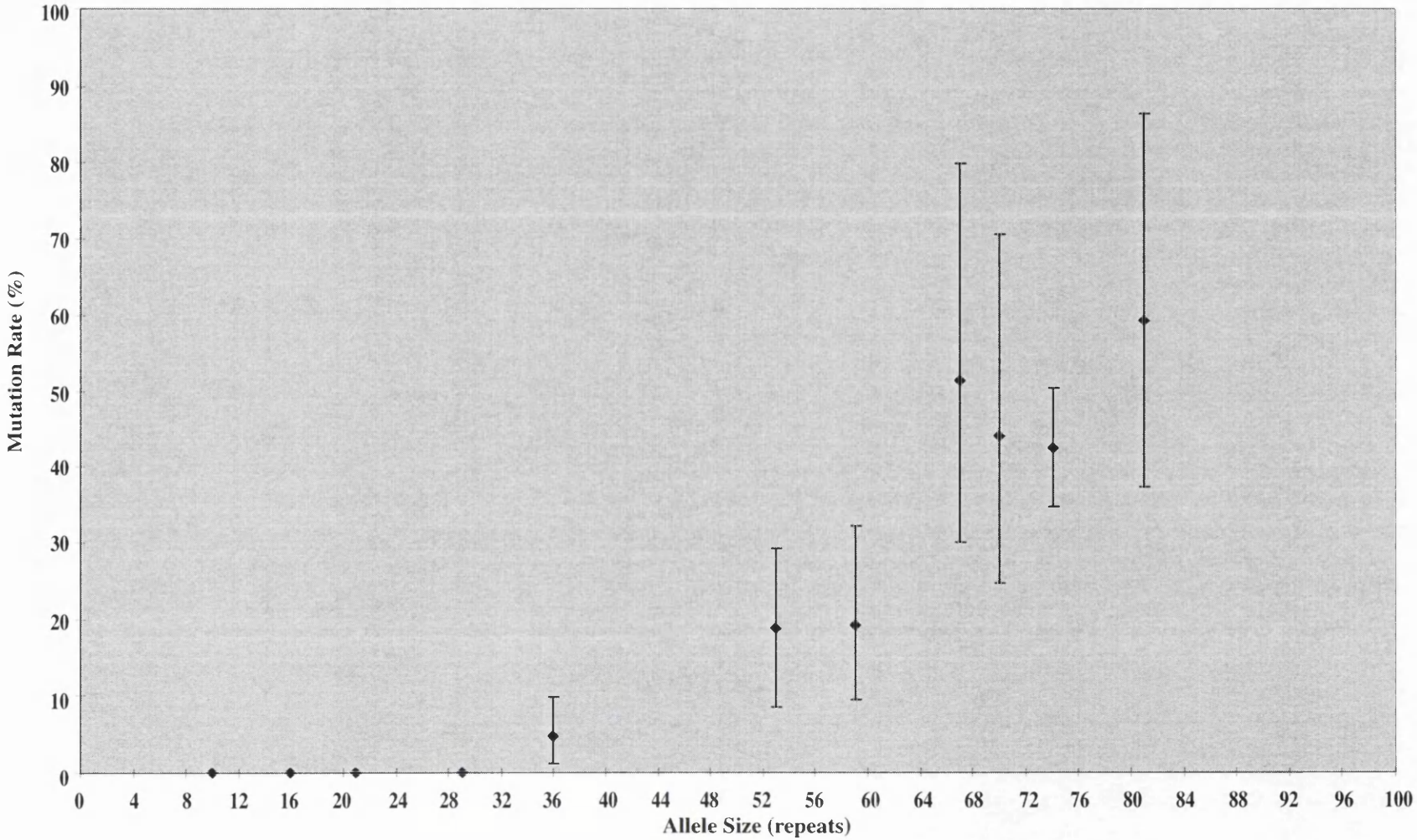

Figure 4.11 Mutation rates of *ERDA1* alleles in the male germline. Displayed are the mutation rates at the single molecule level of *ERDA1* alleles in sperm. Where possible, 95% confidence intervals have been calculated and displayed for each allele. The scatter graph shows increasing mutation rate with increasing allele length, depicted as an exponential curve.

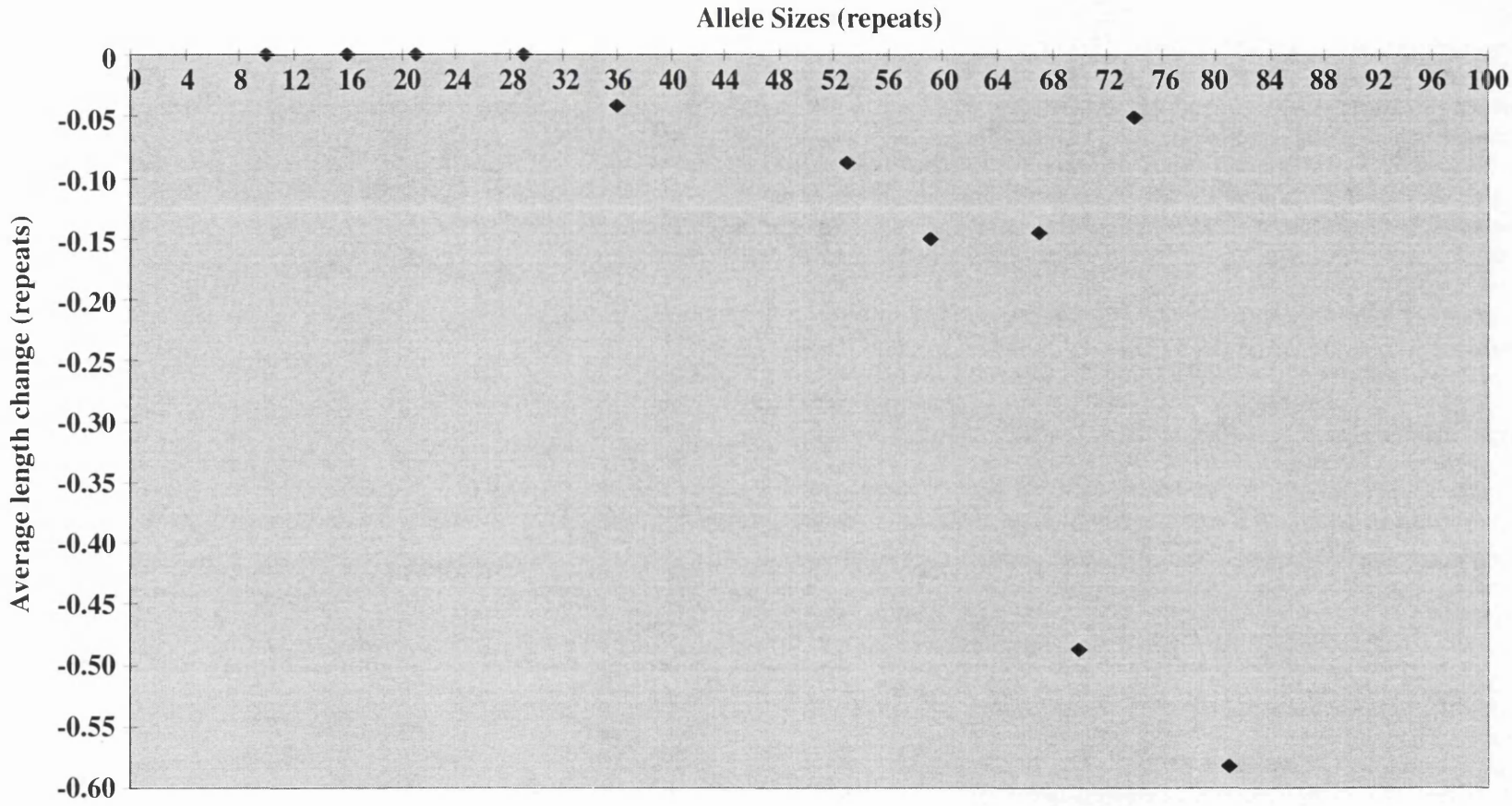

Figure 4.12 The average length change of alleles at the *ERDA1* locus on single molecule DNA isolated from sperm. Displayed are the average length changes at the single molecule level of *ERDA1* alleles in sperm. Alleles up to 29 repeats are stable whereas 36-82 repeats have an increasing mutation rate but biased toward contractions. The 67 repeat of sample GS12 seems to have an average length change within the normal range.

Sample	Progenitor Allele Length	Small length changes							Large length changes			
		No. of reactions (+ve/total)	m	Total molecules analysed	No. mutants	mr	95% CI mr	A.I.c	Total molecules analysed	No. mutants	mr	Combined A.I.c
GS5	15	78/96	1.67	161	0	0	-	0	-	-	-	-
	81*	79/96	1.73	121	99	81.82	38.4-82.7	-0.37	-	-	-	-
GS12	16	49/96	0.71	69	0	0	-	0	7,584	0	0	0
	87*	48/96	0.69	68	60	88.2	83.1-97.3	0.91	7,584	130	1.57	1.502
GS66	11	86/96	2.26	217	0	0	-	0	11,840	0	0	0
	37	79/96	1.73	166	26	15.7	7.5-26.5	-0.024	11,840	0	0	-0.024
GS77	14	98/192	0.78, 0.88	159	0	0	-	0	16,226	0	0	0
	60	83/192	0.3, 0.93	118	61	51.7	37.2-80.4	-0.024	16,226	2	0.02	-0.022
GS124	17	70/96	1.31	125	0	0	-	0	-	-	-	-
	67*	74/96	1.47	103	77	74.75	34.4-77.6	0.31	-	-	-	-
GS130	31	50/96	0.74	71	3	4.2	-	-0.014	4,576	0	0	-0.014
	40	44/96	0.61	59	14	23.7	9.4-47	0.119	4,576	0	0	0.119
GS131	17	64/96	1.1	105	0	0	-	0	4,272	0	0	0
	79*	64/96	1.1	96	83	86.45	79.7-93.3	-1.76	4,272	33	0.73	-1.758
GS133	15	68/94	1.29	121	0	0	-	0	8,640	0	0	0
	26	53/94	0.83	78	3	4	-	-0.038	8,640	0	0	-0.038

Sample	Progenitor Allele Length	Small length changes							Large length changes			
		No. of reactions (+ve/total)	m	Total molecules analysed	No. mutants	mr	95% CI mr	A.l.c	Total molecules analysed	No. mutants	mr	Combined A.l.c
GS148	28	65/96	1.13	108	3	2.8	-	-0.056	5,600	-	-	-
	35	54/96	0.83	80	6	7.5	1.3-17.7	0	5,600	1	0.02	0.002
GS152	17	74/96	1.47	141	0	0	-	0	5,280	0	0	0
	91*	66/96	1.16	129	114	88.37	-	-1.63	5,280	29	0.52	-1.577
GS153	11	72/96	1.39	133	0	0	-	0	7,872	0	0	0
	38	63/96	1.07	102	20	20.6	7.9-34.9	-0.039	7,872	1	0.02	-0.037
GS166	11	87/144	0.93	133	0	0	-	0	1,011	0	0	0
	71*	77/144	0.75	91	70	76.9	42.5-88.7	-0.099	1,011	7	0.52	-0.151
GS167	17	53/96	0.8	77	0	0	-	0	-	-	-	-
	85*	49/96	0.71	62	53	85.47	67.5-87.3	-1	-	-	-	-
GS261	11	53/96	0.8	77	0	0	-	0	5,216	0	0	0
	84*	46/96	0.65	55	48	87.27	78.5-96.1	-0.581	5,216	2	0.04	-0.563
GS273	11	58/96	0.93	89	0	0	-	0	7,552	0	0	0
	76*	65/96	1.13	90	77	85.6	46.2-97.6	-1.712	7,552	20	0.22	-1.733

Table 4.1 Sperm DNA analysis at the *CTG18.1* locus. mr = mutation rate, A.l.c. = average length change, * denotes alleles whose mutation rate was based on percentage frequency rather than poisson calculation. Samples GS5, GS124 and GS167 did not undergo large scale SP-PCR. Combined average length change is the total of small and large length changes.

Sample	Progenitor Allele Length	Small length changes							Large length changes			
		No. of reactions (+ve/total)	m	Total molecules analysed	No. mutants	mr	95% CI mr	A.l.c	Total molecules analysed	No. mutants	mr	Combined A.l.c
GS12	21	53/96	0.8	77	0	0	-	0	5,280	0	0	0
	67	55/96	0.85	82	42	51.2	30-79.8	-0.146	5,280	1	0.02	-0.141
GS20	10	71/96	1.35	129	0	0	-	0	9,024	0	0	0
	59	74/96	1.47	141	27	19.1	9.5-32.1	-0.15	9,024	2	0.02	-0.144
GS94	10	73/96	1.43	137	0	0	-	0	10,304	0	0	0
	36	80/96	1.79	172	8	4.65	1.1-9.8	-0.041	10,304	0	0	-0.041
GS106	16	54/96	0.83	79	0	0	-	0	5,376	0	0	0
	70	55/96	0.85	82	36	44	24.6-70.7	-0.488	5,376	0	0	-0.488
GS110	10	78/96	1.67	161	0	0	-	0	10,688	0	0	0
	74*	93/96	3.47	160	68	42.5	34.8-50.2	-0.051	10,688	0	0	-0.051
GS117	10	90/96	2.77	266	0	0	-	0	18,304	0	0	0
	53	91/96	2.95	284	53	18.7	8.5-29.1	-0.088	18,304	0	0	-0.088
GS174	16	68/96	1.23	118	0	0	-	0	7,456	0	0	0
	29	64/96	1.1	106	0	0	-	0	7,456	1	0.01	0.004
GS306	10	49/96	0.71	69	0	0	-	0	5,696	0	0	0
	81	63/96	1.07	103	61	59.2	37.2-86.4	-0.582	5,696	0	0	-0.582

Table 4.2 Sperm DNA analysis at the *ERDA1* locus. mr = mutation rate, A.l.c. = average length change, * denotes alleles whose mutation rate was based on percentage frequency rather than poisson calculation. Combined average length change is the total small and large length changes.

Chapter 5

Other factors influencing repeat stability

5.1 Somatic mosaicism at the *CTG18.1* locus

5.1.1 Introduction

One of the simplest tissues to obtain from patients is blood and hence SP-PCR has been used to observe repeat behaviour in tissues other than the germline (i.e. in the soma). In the case of myotonic dystrophy, extensive SP-PCR analysis has been carried out on blood and muscle, either because of their ease of extraction or their association with the disease phenotype. Muscle is particularly prone to expansion, but blood is more stable and often used for diagnostic purposes. However, it has been noted that over a period of time, the repeat distribution in the blood can also "shift" to a larger average size with a greater spread of alleles (Martorell *et al.*, 1998) and hence the age of the individual and the size of the repeat can make it increasingly difficult to predict the progenitor. This has consequences for diagnostic purposes where it may seem there are apparent contractions in transmission of the repeat, and correlations between genotyping and age of onset which could suggest that symptoms could develop earlier than would be the case.

It is known that the *DM-1* repeat shows contrasting behaviour in cultured lymphoblastoid cell lines to the somatic tissue it is derived from (Ashizawa *et al.*, 1996). CEPH DNA is extracted from transformed lymphoblastoid cell lines established from the lymphoblasts of individuals from large families. Cell lines established with clones from *DM-1* patients demonstrated that the repeat does mutate, but not at a high rate (Ashizawa *et al.*, 1996). SP-PCR showed that these mutations are biased toward contractions, from a progenitor of 230 repeats at a rate of 40% for small length changes and 1% for large length changes, increasing to 75 and 20% respectively with further doublings. A progenitor of 1100 repeats had a mutation rate of 70% for small length changes and 10% for large length changes, also biased toward contractions. The first publication of the *CTG18.1* locus analysed BPAD families for expanded alleles as well as CEPH DNA to ascertain whether instability at this locus could contribute to the phenotype (Breschel *et al.*, 1997). Expanded alleles were detected in CEPH and BPAD DNA as well as unaffected individuals in disease families. Therefore, *CTG18.1* was discounted as a BPAD mutation. It was reported however, that large *CTG18.1* expanded alleles were detected

in two CEPH families, 1344 and 1420. Therefore in order to observe expanded repeat behaviour in DNA derived from somatic tissues, CEPH DNA from families 1332, 1344 and 1420 and blood DNA from DM-1 patients were analysed by SP-PCR at the *CTG18.1* locus.

5.1.2 Results

In order to observe somatic behaviour at the *CTG18.1* locus and compare to the male germline, blood DNA available from 50 *DM-1* patients were typed (Appendix 1) and five individuals were identified to have *CTG18.1* expansions. The range of alleles detected was 11-89 repeats. All five samples underwent preliminary SP-PCR analysis at *CTG18.1* and three are displayed in Figure 5.1. Samples 101-4004 and 101-3056 are relatively stable having upper alleles of 51 and 69 repeats respectively, with a mutation frequency of approximately 16-18%. Sample 152-3021 has a larger allele of 89 repeats and demonstrates an increased mutation frequency of 37% but with expansions, rather than contractions as observed in the male germline. The distribution of alleles in the blood is similar to that seen at the *DM-1* locus in patients, in that there is a defined lower boundary and a bias toward expansions, though not to the same scale (Monckton *et al.*, 1995, Martorell *et al.*, 1998).

The CEPH DNA of the Fathers (F-01) and Mothers (M-02) of families 1332, 1334, 1344,1420 and 1424 were typed by PCR at the *CTG18.1* locus. Alleles ranged from 11-330 repeats (Appendix 1). Alleles up to 118 repeats were detectable by fluorescent PCR and the expanded allele of 330 repeats was estimated by PCR Southern blot. Preliminary SP-PCR at the *CTG18.1* locus on samples M-02 1332 (11, 52) and M-02 1344 (17, 118) and large scale SP-PCR on sample F-01 1420 (15, 330) are displayed in Figure 5.2. The 52 repeat allele is relatively stable, demonstrating approximately 2% large length changes, the 118 repeat allele is demonstrating 8.7% large length changes, and the 330 repeat allele is mutating at a frequency of 11%, sometimes into the normal range. This contrasts with the observed increased instability in the blood DNA from DM-1 patients for similarly sized alleles and there is an upper boundary for mutations as opposed to a lower boundary.

5.1.3 Discussion

A limited number of blood DNA samples with *CTG18.1* expansions were available to analyse tissue-specific differences, but a behaviour change can be detected. Alleles at the *CTG18.1* locus have a tendency toward contraction in the male germline but to expand in the blood. SP-PCR data on triplet repeat sequences indicates trinucleotides are more stable in blood than the germline, but there is a lack of published data for an equivalent repeat range as that examined in this section. For example, extensive SP-PCR has been carried out on blood DNA from *DM-1* patients (Monckton *et al.*, 1995, Martorell *et al.*, 1998), demonstrating that although the repeats are more stable in this tissue than others such as sperm and muscle, there are frequent mutations that are biased toward expansions. The published data available for *DM-1* alleles of the 50-100 repeat range suggests there are mutations ranging from approximately 30-100 repeats (Monckton *et al.*, 1995). Unpublished SP-PCR data does suggest that alleles of 60-90 repeats have increasing mutation rates with the occasional expansion up to 210 repeats for larger alleles and one sample with a 118 repeat allele had a range of 98-275 repeats (Hogg and Monckton, unpublished). Therefore the rate of mutation at the *CTG18.1* locus in sample 152-3021 (27, 89) is subtle compared to the behaviour observed at the *DM-1* locus with similarly sized alleles. The range of ages for the blood samples is 20-64 years of age and the youngest individual sampled was 152-3021, indicating that if there is an age effect, it is subtle compared to *DM-1*. It would be interesting to see if there is a time effect at the *CTG18.1* locus as in the case of *DM-1* and 152-3021 would be a good candidate for that study, due to the size of the allele and the age of the individual. This may also assist in models of *DM-1* repeat behaviour where the alleles are so unstable that it is difficult to define a simple model.

CEPH DNA is available in order to provide large numbers of transmissions within families in which to study human genetics. In the case of the *CTG18.1* repeat, CEPH families were screened in order to establish that normal individuals do not possess expanded alleles (Breschel *et al.*, 1997). However, various expanded alleles were discovered in healthy individuals with no history of serious illness. It was not stated how large these alleles were and so these individuals were typed at the *CTG18.1* locus. Sample F-01 1420 (15, 330) had the largest allele detected by PCR which was evident only on a PCR Southern blot and could not be detected using fluorescent primers. Therefore as this allele was the largest available at the *CTG18.1* locus, it was subjected to SP-PCR. The data seems to indicate that large alleles are relatively stable in cultured lymphoblastoid cell lines and show contrasting behaviour to the tissue they originate from. From the *DM-1* patient blood data, it is known that the *CTG18.1* repeat is likely to expand rather than contract with increasing allele length in blood. It is also known that the *DM-1* repeat is biased toward expansions in the blood as

demonstrated by SP-PCR (Monckton *et al.*, 1995, Martorell *et al.*, 1998). However it is also known that the *DM-1* repeat also shows contrasting behaviour in cultured lymphoblastoid cell lines (Ashizawa *et al.*, 1996). Cell lines established from a clone from *DM-1* patients demonstrated that the repeat does mutate, but not at a high rate, and SP-PCR showed that these mutations are biased toward contractions from progenitors of 230 and 1100 repeats. Sample F-01 1420 also showed that this large allele of 330 repeats at the *CTG18.1* locus is contracting and no expansions were detected, although only a total of 758 molecules were analysed. *DM-1* alleles do expand occasionally in these cell lines, whereas the expanded *CTG18.1* allele was seen to be more stable in regard to expansions. Therefore it seems that large, expanded alleles will generally contract in cultured lymphoblastoid cell lines but the flanking sequence of the repeat determines the subtle mutations that occur and can only be detected by SP-PCR. The *DM-1* locus is more unstable than the *CTG18.1* locus and so still demonstrates some expansions from a 230 repeat allele. The *CTG18.1* locus is not as prone to gross expansions from equivalently sized alleles and so this 330 repeat is not expanding though it might be expected to, given its length.

5.2 Instability at the *CTG18.1* locus in MMR deficient tissues

2.5.1 Introduction

Since the discovery of the DNA mismatch repair gene mutations giving rise to HNPCC, much work has focused on microsatellite instability at many loci distributed over the entire genome. Various approaches have been taken in order to select which markers may be more informative or sensitive to the RER+ phenotype, either based on previous MIN+ observations, marker location in respect to the MMR genes, or the availability of markers within the research group. Most observations have been with dinucleotide (CA) markers due to their availability in the genome. Microsatellite instability has also been investigated in various types of ovarian cancer, although the overall incidence is low compared to that seen in HNPCC tumour DNA. MIN+ is usually associated with the less severe forms and early FIGO stages of ovarian cancer (Park *et al.*, 1995, Pieretti *et al.*, 1995, Tangir *et al.*, 1996), although it is not widespread in the genome, whereas in HNPCC MIN+ is much more common at a wide range of markers (Aaltonen *et al.*, 1993, Ionov *et al.*, 1993, Lindblom *et al.*, 1993, Thibodeau *et al.*, 1993, Aaltonen *et al.*, 1994, Pedroni *et al.*, 1999). Cases of ovarian cancer where the tumour has demonstrated instability at more than one marker are therefore thought to have derived from individuals within HNPCC families (Arzimanoglou *et al.*, 1996, Phillips *et al.*, 1996) and

MMR mutations have been detected in some of these tumours. Of great importance is to characterise the underlying germline MMR mutation as it has been noted that *hMLH1* mutations seem to confer a higher rate of microsatellite instability than *hMSH2* mutations (Lamberti *et al.*, 1999, Wu *et al.*, 1999). Therefore we sought to test if microsatellite instability could be detected at the *CTG18.1* locus in ovarian tumours and whether this instability could be replicated in HNPCC tumours with a known MMR deficiency.

2.5.2 Results

DNA was available from an ovarian carcinoma cell line, A2780 and cisplatin resistant derivatives of this cell line, known as MCP1-9 (Anthoney *et al.*, 1996). A2780 is known to be proficient in *hMLH1, hMSH2, hPMS2* and *hMSH6*, whereas the MCP derivatives are deficient in *hMLH1* and *hPMS2* (Brown *et al.*, 1997). MCP1 and 8 are partially methylated at four *Hpa*II sites in the *hMLH1* promoter, whereas MCP3, 5 and 7 are fully methylated (Strathdee *et al.*, 1999). Cell line LoVo is deficient in both *hMLH1* and *hMSH2* and AN3CA is an endometrial cell line deficient in *hMLH1* (Umar *et al.*, 1994). In order to test if any deficiencies in the MMR genes could alter the stability of trinucleotide repeats, these cell lines were typed at the *CTG18.1, ERDA1* and *DM-1* loci. Alleles ranged from 10-28, 9-20 and 5-13 repeats respectively (Appendix 1). The derived cell lines MCP1-9 were stable at the *DM-1* and *ERDA1* loci, but the upper *CTG18.1* allele of 27 repeats mutated by –1 and +1 repeats at the *CTG18.1* locus respective to the parental cell line A2780. This suggests that the alleles at the *DM-1* and *ERDA1* loci are too small to be affected by defects in mismatch repair, but that the larger allele at the *CTG18.1* locus is large enough to mutate due to errors in replication and hence would avoid correction.

Blood and ovarian tumour DNA was also available from 15 individuals who were typed at *CTG18.1, ERDA1* and *DM-1*. Alleles ranged from 11-64, 10-68 and 5-26 repeats respectively (Appendix 1). Given the sensitivity already noted in the cell lines at the *CTG18.1* locus, blood sample LC (11, 64) underwent preliminary SP-PCR at the *CTG18.1* locus. No increased instability was observed from the blood DNA samples analysed in Section 5.1. SP-PCR on the tumour DNA from this individual proved difficult and no definite conclusions could be drawn on instability due to PCR inhibitors that may be a result of the paraffin fixing process.

DNA extracted from four ovarian tumours were typed and alleles ranged from 11-106, 10-60 and 5-13 repeats at *CTG18.1, ERDA1* and *DM-1* respectively (Appendix 1). Sample OT481 (17, 106) underwent preliminary SP-PCR at the *CTG18.1* locus and demonstrated high levels

of instability of the upper allele, heavily biased toward expansions (Figure 5.3) although the majority of alleles are close in size to the progenitor, which is demonstrated by fluorescent PCR (Figure 5.9). There is a definite lower boundary and though alleles have increased to over 600 repeats in length, the majority of upper alleles are between 106 and 170 repeats. There is no additional information such as age of individual, stage of tumour or germline MMR mutations which would help to assess if the behaviour observed is a response to defects in mismatch repair or other factors. In order to test if this microsatellite instability is widespread, other trinucleotide repeat loci were investigated. Normal alleles at loci *DM-1* (5, 13), *ERDA1* (21, 60), *SBMA* (17, 25) and *SCA1* (28) were stable in this tumour sample (Figure 5.4). Therefore, this gross instability appears to be specific to the expanded allele at the *CTG18.1* locus.

The mismatch repair status of OT481 is unknown, so in order to discover if this gross microsatellite instability could be replicated in tumours known to have defective mismatch repair status, loci *CTG18.1* and *ERDA1* were investigated to see if expanded alleles would behave differently in HNPCC tissue, either tumour or non-tumour compared to the somatic behaviour already observed in blood (Section 5.1). Blood DNA was available from a total of 83 HNPCC individuals from 53 families and 162 chromosomes were typed at both *CTG18.1* and *ERDA1*. The ranges of alleles were 11-93 and 7-86 repeats respectively (Appendix 1). In order to screen the blood DNA samples for expanded alleles demonstrating instability without using SP-PCR, typing PCR products at the *CTG18.1* locus were electrophoresed through a 1.75% agarose gel, subjected to a Southern blot and radioactively hybridised. Figure 5.5 displays the Southern blot of the PCR typing at the *CTG18.1* locus on blood DNA from HNPCC individuals. Two samples, EE and her mother FT, have both inherited the same germline 2bp deletion in exon 19 of *hMLH1* and an expanded *CTG18.1* allele of 93 repeats. These two samples have produced a smear on the Southern blot in addition to the upper allele that can be detected at approximately 100 repeats. This indicates a varied population of repeat sizes in these samples and so the more sensitive technique of SP-PCR was applied at this locus in these samples and a third, BeMa, which also has an expanded allele of 81 repeats (not shown).

To establish the mutation rate of the 93 repeat allele in normal tissue from HNPCC samples EE and FT and the 81 repeat allele in BeMa, blood DNA underwent preliminary SP-PCR at the *CTG18.1* locus (Figure 5.6). The 93 repeat alleles of FT and EE clearly show dramatic, expansion biased instability which can also be detected by fluorescent PCR though not to the same extent (Figure 5.9). The upper allele of FT has expanded to greater than 770 repeats at a high frequency with no definite lower boundary. The upper allele of daughter EE is also expanding to greater than 770 repeats but the majority of alleles are between 93 and

approximately 270 repeats in size, with a definite lower boundary at the progenitor allele of 93 repeats. As both individuals inherited the same MMR mutation it is assumed that this difference is due to the age effect. FT was 83 years old at time of sampling and her daughter EE was 54. The 81 repeat allele of BeMa however is much more stable both in SP-PCR (Figure 5.6) and fluorescent PCR (Figure 5.9). Unfortunately there is no information available about the age or inherited MMR mutation for BeMa, although the individual is female, as in the cases of FT and EE. BeMa has a similarly sized expansion allele at *CTG18.1* but it is much more stable than the upper alleles of FT and EE. However, this allele is still more unstable than similarly sized alleles in the blood of *DM-1* patients (Section 5.1).

The extreme behaviour of the expanded *CTG18.1* allele in samples FT and EE compared to the blood DNA samples analysed from *DM-1* patients (Section 5.1) and HNPCC sample BeMa suggests that it is a response to a specific MMR mutation. The blood DNA is heterozygous for the 2bp deletion and so the tumour would be predicted to have accumulated another mutation in *hMLH1* in the colon tissue. It was therefore expected that the expanded *CTG18.1* allele would be relatively stable in blood in order to compare against any increased instability observed in tumour DNA. However, the preliminary SP-PCR suggests that some additional mechanism may be operating to contribute to this gross instability. To determine if other trinucleotide repeats were also being affected by this additional mechanism, the blood DNA of FT and EE underwent SP-PCR at the *DM-1, ERDA1, SBMA* and *SCA1* loci (Figures 5.7 and 5.8). Both samples had normal alleles at all four loci which were stable.

Tumour DNA was available for EE and so it also underwent preliminary SP-PCR to assess the behaviour of the *CTG18.1* locus in response to total deficiency in *hMLH1* (Figure 5.10). Only the lower allele could be detected in the tumour DNA and this seemed unstable. There was very little DNA in the sample as there were only intermittent positive reactions for the lower allele, but using undigested, and hence more concentrated DNA instead of *Hin*dIII digested DNA did not increase the number of positive reactions. This suggests that the DNA could have been damaged when the tumour was set in the paraffin block and so PCR would not work as well. The other possibility is that the repeat has expanded outside the limits of detection by PCR. In order to visualise PCR products from the upper allele that could not be resolved on the SP-PCR blot, the 8 reactions for either digested or undigested underwent a second round fluorescent PCR and the products electrophoresed through a 4.75% polyacrylamide gel and analysed with GeneScan 3.1 software (Figure 5.11). The upper allele was not detected and so it was concluded that there was not enough DNA from the tumour to analyse the behaviour of the expanded *CTG18.1* allele in response to total deficiency of *hMLH1*.

Expanded *ERDA1* alleles were also available to analyse if this locus could demonstrate the same type of instability seen at *CTG18.1*. Preliminary SP-PCR was carried out at the *ERDA1* locus (Section 2.3) on HNPCC blood samples JC (21, 62) and NS (10, 75) (Figure 5.12). JC has inherited a G to A transition in exon 19 of *hMLH1* creating a premature STOP codon and NS has inherited a 1bp deletion in exon 6 of *hMSH2*. Both alleles of each sample are very stable at the *ERDA1* locus in blood. From the male germline data in Section 4 it is known that the *ERDA1* locus is much more stable than *CTG18.1*. Therefore it is not surprising that it is also very stable in blood and unlikely to produce mutations that would evade correction by the mismatch repair machinery. Therefore it would still appear stable.

Colon tumour DNA available from HNPCC individual BaF, who has inherited a 2bp insertion in exon 1 of *hMLH1* and an *ERDA1* expansion of 60 repeats, demonstrates stability at the *ERDA1* locus at both the normal and expanded alleles (Figure 5.10). There is an extra band evident in the sample, but this product is often detected in *ERDA1* SP-PCR, varying in size between the two alleles, although it is not contamination and thought to be an artefact. The brother of BaF, BF also inherited the same HNPCC germline mutation and expanded *ERDA1* allele and also demonstrated stability in the blood (data not shown). Tumour tissue from BF was fixed in a paraffin block, the DNA extracted and subjected to SP-PCR at the *ERDA1* locus. Although both alleles could be detected with typing primers ERDA-C/DR on an agarose gel, neither allele was observed with SP-PCR primers ERDA-A/BR. In order to visualise the PCR products that could not be resolved on the SP-PCR blot, the 8 reactions of the most concentrated input DNA underwent a second round fluorescent PCR and the products electrophoresed through a 4.75% polyacrylamide gel and analysed with GeneScan 3.1 software (Figure 5.11). Both alleles were detected, but artefacts can also be seen. However, it does seem that this tumour DNA is also stable at this locus for both the expanded and normal alleles, as in tumour sample BaF.

2.5.3 Discussion

The cell line DNA all possessed small alleles at the *DM-1, CTG18.1* and *ERDA1* loci and given the observations that only expanded alleles seem to demonstrate instability, SP-PCR was not pursued in these samples. However, it was noted that the *hMLH1* deficient cell lines MCP1-9 derived from a MMR proficient cell line A2780, are unstable for the 27 repeat at the *CTG18.1* locus. Of the seven cell lines known to be *MLH1* deficient, 4/7 gained a repeat, 1/7 lost a repeat and 2/7 remained stable. Though information on the status of *hMLH1, hMSH2, hPMS2* and *hMSH6* is available, it does not directly correlate with the gain or loss of a repeat.

It cannot be ignored however that the 27 repeat is demonstrating a higher frequency of mutations than would be expected if the cell lines were *hMLH1* proficient.

Given the limited number of ovarian tumour DNA samples available for SP-PCR analysis, it was fortunate to discover a sample with an expanded allele at the *CTG18.1* locus, as the frequency of expanded alleles is 3%. The expanded allele of 106 repeats in sample OT481 is demonstrating a higher mutation rate compared to the blood DNA from *DM-1* patients in Section 5.1. There is no data available regarding the MMR status, age of individual or type of ovarian tumour which could give some clues as to why this expanded allele is demonstrating gross instability at the *CTG18.1* locus. It could be this individual has inherited a *hMLH1* mutation which is most likely to contribute to microsatellite instability, that she could have an early age of onset given the definite lower boundary of the expanded allele, or that the tumour itself could have been in the early FIGO stages as they have the greatest frequency of microsatellite instability. SP-PCR at other trinucleotide loci *DM-1, ERDA1, SBMA* and *SCA1* confirmed that the observed microsatellite instability was specific to *CTG18.1*, although all loci had normal alleles which have reduced mutation rates and hence it would be improbable that similar behaviour would be evident.

Microsatellite instability was detected in HNPCC tumour samples before the underlying MMR mutations were identified (Aaltonen *et al.*, 1993). It has been postulated that 75-100% of HNPCC tumours demonstrate microsatellite instability whereby the size of alleles differs between normal and cancerous tissue from the same individual (Aaltonen *et al.*, 1993), (Aaltonen *et al.*, 1994, Lindblom *et al.*, 1993) compared to sporadic colon tumours that show a low frequency of the MIN+ genotype (Pedroni *et al.*, 1999). Observations seem to suggest that mutations in the *hMLH1* gene confer a higher rate of microsatellite instability than other MMR genes (Lamberti *et al.*, 1999, Wu *et al.*, 1999), and 90% of sporadic tumours with high MIN+ status have loss of expression or hypermethylation of *hMLH1* (Cunningham *et al.*, 1998, Thibodeau *et al.*, 1998). In order to observe microsatellite instability at the *CTG18.1* and *ERDA1* loci, HNPCC blood samples were screened and expanded alleles were identified. Preliminary SP-PCR uncovered extreme instability of a 93 repeat allele at the *CTG18.1* locus in two members of a HNPCC family with a 2bp deletion in exon 19 of the *hMLH1* gene. This instability in the normal tissue in response to a heterozygous MMR mutation has been detected before at di and mononucleotide markers in the lymphoblasts and colon of two siblings who have inherited the same *hPMS2* mutation (Parsons *et al.*, 1995). There had been no subsequent somatic mutation of the normal *hPMS2* gene or of other MMR genes, so it was suggested that this mutation was creating a dominant negative effect on the protein. In this case, the final six amino acids of *hMLH1* preceding the STOP codon have been altered to a Serine and Leucine before a premature STOP. These final six amino acids are 100%

conserved in the human, rat, yeast and *Drosophila* homologues and only one amino acid change is observed in the *Arabidopsis thaliana MHL1* gene (Figure 5.13) (Prolla *et al.*, 1994, Kolodner *et al.*, 1995, McKee 1998, Geeta Vani *et al.*, 1999, Jean *et al.*, 1999). The fact that the extreme carboxy terminus of this gene is so highly conserved suggests these final amino acids are essential for the function or location of *MLH1*, although it is not known whether this region is part of a domain that could be responsible for protein interactions. The gross instability observed in blood samples FT and EE at the *CTG18.1* locus therefore suggest a possible dominant negative mutation in the *hMLH1* gene. It could be that the change in the amino acid sequence of the extreme carboxy terminus would allow the mutated protein to interact with the normal *hMLH1* protein or proteins from the MMR pathway and hence reduce overall function. Alternatively, the mutated *hMLH1* gene could not be directly responsible for the observation at the *CTG18.1* locus and inherited haplotypes in the flanking sequence or mutations in other genes could be major contributors. Ideally, we would need to analyse other members of this family to assess the expanded *CTG18.1* repeat in individuals who do not possess the *hMLH1* mutation. Only by analysing more blood DNA samples from normal individuals with an expanded *CTG18.1* repeat can this gross instability be attributed to defects in mismatch repair.

```
rMLH1       LANLPDLCKVFERC
AtMLH1      VASLEKLYKIFERC
hMLH1       LANLPDLYKVFERC
dMLH1       LTNLPTLYKVFERC
yMLH1       IANLPDLYKVFERC

EE/FT       LANLPDLYSL----
```

Figure 5.13 Comparison of *hMLH1* C-terminus amino acid sequence. Alignment of the last 14 amino acids of the *MLH1* gene from rat (Geeta Vani *et al.*, 1999), *arabidopsis thaliana* (Jean *et al.*, 1999), human (Kolodner *et al.*, 1995), *Drospohila* (McKee 1998) and yeast (Prolla *et al.*, 1994). The last four amino acids are 100% conserved between all species, suggesting they are extremely important.

Both samples are displaying expanding repeat behaviour in response to increasing age of individual as seen in other triplet repeats such as *DM-1* (Monckton *et al.*, 1995, Martorell *et al.*, 1998). Sample FT has a less definite lower boundary than EE and a greater frequency of very large alleles. This suggests therefore that age dependent expansion is a feature of large CTG alleles independent of the flanking sequence. The extreme microsatellite instability observed for the ovarian tumour OT481 and HNPCC tissues has been noted for the *DM-1* CTG repeat in tumourous tissue from *DM-1* patients. A Japanese woman was typed to have approximately 1,500 repeats in the blood, but over 5,000 repeats in the ovarian cancer (Kinoshita *et al.*, 1997) and three *DM-1* patients with various cancers had increases of up to 3,000 repeats in the tumour compared to DNA from the non-tumour tissue (Jinnai *et al.*, 1999).

Further investigations into possible instability at other trinucleotide repeat loci in blood DNA samples FT and EE revealed that normal alleles were stable and the instability observed is specific to the expanded allele at the *CTG18.1* locus. Therefore the small alleles at the *CTG18.1*, *DM-1*, *ERDA1*, *SBMA* and *SCA1* loci are easily replicated within the blood DNA with no errors incurred. SP-PCR at the *CTG18.1* locus in the tumour DNA of EE suggests the second "hit" in the MMR machinery that subsequently triggered tumour formation is also failing to correct errors introduced into the small allele. The large allele of 93 repeats could not be detected which suggests there was low quality DNA from the tumour or that the upper allele has expanded outside the limits of PCR detection. If more concentrated DNA was available from this tumour, then the *DM-1, ERDA1, SBMA* and *SCA1* loci could be analysed for any differences to the blood, or to answer the question of the size of the upper allele, possibly by restriction digestion and a Southern blot.

The lack of information regarding the germline mutation inherited by individual BeMa means it is not possible to comment on the relative stability of the expanded allele compared to FT and EE, although the blood results would provide a basis for comparisons with tumour DNA. Due to the prevalence of different MMR mutations giving rise to HNPCC, it is assumed that HNPCC blood sample BeMa does not have the same MMR mutation. If BeMa does have the same mutation however, it is possible that this individual is young and there could be an age effect which would explain the difference to FT and EE.

The *ERDA1* locus was also investigated to observe any microsatellite instability in HNPCC tissue. Expanded alleles were detected and the blood and tumour DNA from two siblings, BaF (10, 60) and BF (10, 60) who inherited a 2bp insertion in exon 1 of *hMLH1*, were subjected to SP-PCR. Blood DNA was stable for each allele at this locus in both individuals and investigation of tumour DNA also found the locus is stable. The *ERDA1* locus also

proved very stable in blood samples JC (21, 62) and NS (10, 75). JC inherited a G to A transition in exon 19 of *hMLH1* creating a premature STOP codon and NS inherited a 1bp deletion in exon 6 of *hMSH2*. The mutation JC inherited is further upstream in *hMLH1* than the premature STOP codon inherited by individuals FT and EE, suggesting that either the *ERDA1* locus does not have a sufficient level of instability to be affected by defective *hMLH1* or that the change of amino acids is more important than the STOP codon in altering *hMLH1* function. In future experiments, detection of *ERDA1* alleles in fixed tissue DNA could be improved by carrying out a first round SP-PCR followed by a second round fluorescent PCR. In tumour sample BF, *ERDA1* alleles could not be detected from first round SP-PCR, but were clearly observed after second round fluorescent PCR and GeneScan analysis. This confirmed that this sample was also stable.

Many ovarian tumour and blood samples from the same patients need to be screened in order to verify the results depicted in this section. Testing for MMR mutations and other features of ovarian cancer genetics such as LOH and mutations in genes such as *KRAS, BRCA-1* and *p53* and classification of the FIGO stages and type of tumour would help to define the different mechanisms of mutation occurring. There also seems to be no continuity between sampling and analysis of markers in ovarian cancer or HNPCC, and although there are indications that certain types of tumour are more likely to be MIN+ than others, the crude way of determining microsatellite instability with larger quantities of DNA may not detect the type of instability seen in samples OT481, FT and EE. Dinucleotide repeats also seem to be stable which is reflected in the range of alleles at these markers. Therefore they are much less likely to mutate and may not be ideal candidates for observing microsatellite instability.

5.3 Repeat behaviour at the *CTG18.1* locus in response to Radiotherapy treatment

5.3.1 Introduction

Analysis of samples in the previous sections has primarily been concerned with the natural behaviour at the *CTG18.1* or *ERDA1* loci in the male germline, and how the alleles at the *CTG18.1* locus exhibit different behaviour in somatic tissue and in response to deficient mismatch repair. These differences are due to the tissue or inherited genomic deficiencies in the individual. However, it has been suggested that influences such as chemical or ionising radiation can also change the status of the human genome. One example of chemical radiation affecting DNA is the fallout from the Chernobyl disaster. There has been a general increase in the incidence of cancer in families from areas such as Belarus and though elevated mutation rates have been observed for minisatellites in the germline of these sufferers (Dubrova *et al.*, 1997), there has been no significant microsatellite instability detected in the tumours (Nikiforov *et al.*, 1998). With regard to ionising radiation, there was no significant difference in the mutation rate of a minisatellite, MS205 in sperm before and after radiotherapy, detectable by SP-PCR (Armour *et al.*, 1999). Due to the sensitivity of the *CTG18.1* locus in response to defective MMR, this locus was tested to see if the stability of a microsatellite could be affected by exposure to radiation.

5.3.2 Results

Sperm DNA samples were obtained from several patients before and after radiotherapy treatment for testicular cancer. Typing PCRs were carried out at the *DM-1*, *CTG18.1* and *ERDA1* loci with ranges of 5-34, 11-31 and 10-65 repeats respectively (Appendix 1). Sperm DNA samples taken before and 65 days following treatment of individual XRT1970 (5, 34) (11, 31) (22, 22) underwent single molecule SP-PCR at both the *CTG18.1* and *DM-1* loci, followed by a second round fluorescent PCR and the products electrophoresed through a 4.75% polyacrylamide gel and analysed by GeneScan 3.1 software. The preliminary SP-PCR at each locus does not indicate any gross differences between the samples (Figure 5.14). The frequencies of the larger allele at each locus following second round fluorescent PCR analysis are displayed in Table 5.1 and Figure 5.15. The *CTG18.1* locus was stable before and after treatment with a slight increase in the number of progenitor alleles after treatment, from 94.5 to 95.3% and a tighter distribution of mutations from 25-33 repeats to 30-33 repeats. The *DM-1* locus is known to be more unstable and though the difference in progenitor alleles at these two loci is only 3 repeats, there is approximately a 40% difference in progenitor

frequency. The progenitor allele of 34 repeats increases from 41.7 to 47.8% at the *DM-1* locus after treatment and the distribution of alleles changes from 27-42 repeats to 28-39 repeats. These slight differences between the samples before and after treatment are most likely a result of the time difference between collection of samples or limits of the number of molecules analysed for each locus.

5.3.3 Discussion

Though only one sperm sample was analysed in detail at two trinucleotide repeat loci, there does not seem to be an effect of radiation on their behaviour. It would be assumed that the radiation would induce a mutation in genes that would either leave mutations in these microsatellites unchecked or introduce a greater frequency of errors into the length of the repeat. In fact the trinucleotides seem to be stable and the slight differences between the two samples at either locus is due to the small number of molecules analysed or just stochastic variation. Based on single molecule observations at the *DM-1* locus in the male germline (Appendix 2), the *DM-1* 34 repeat allele has a mutation rate and average length change expected for a 35-36 repeat allele. This mutation rate and average length change are slightly elevated from that displayed in the *DM-1* single molecule data but likely to be a result of small numbers of samples analysed in that range. The behaviour of the 31 repeat allele at the *CTG18.1* locus is much more subtle, which is to be expected given the smaller allele length and the stability of this locus compared to *DM-1*. A Chi-squared test on the positive reactions found no evidence for meiotic drive between the upper and lower alleles at either locus.
The limited data documented here in combination with studies on individuals born to parents who had paediatric cancer found no increased risk of genetic disease (Byrne *et al.*, 1998). Therefore this suggests there are no adverse effects of radiotherapy on the male germline. Further studies at microsatellite and minisatellite loci would help to advise male patients of the risks of having children after radiotherapy treatment.

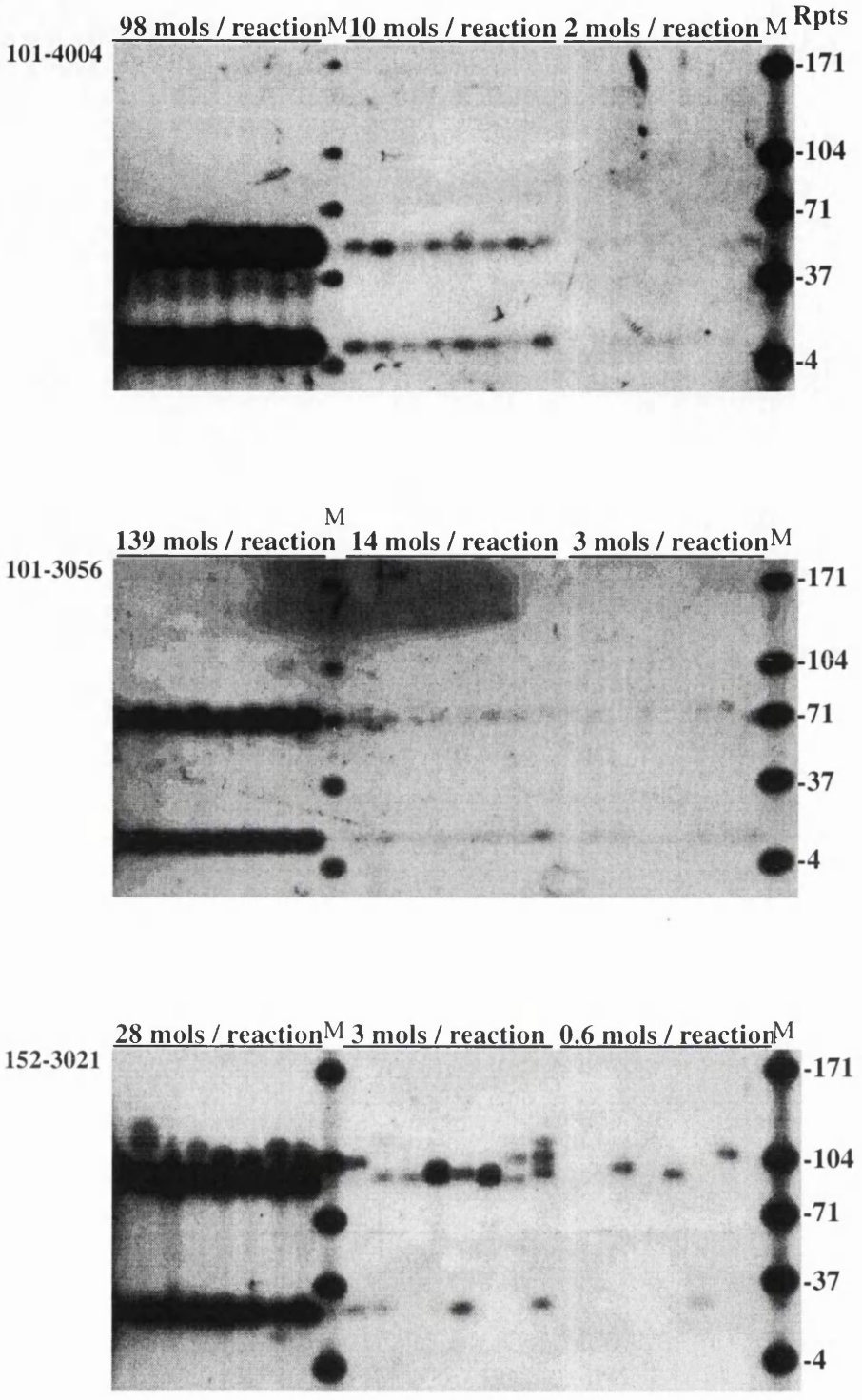

Figure 5.1 Preliminary SP-PCR on blood DNA at the *CTG18.1* locus. Preliminary SP-PCR at the *CTG18.1* locus was carried out on blood DNA extracted from *DM-1* patients. Displayed are samples 101-4004 (11,51), 101-3056 (15,69) and 152-3021 (27,89) with respective expanded *DM*-1 alleles of 800, 500 and 500 repeats. Amplisize/2.5Kb marker bands are labelled as equivalent sizes in repeats. Mols = molecules. There is an increasing mutation rate with increasing allele size and a defined lower boundary below which mutations do not occur. Mutations are biased toward expansions but at a lower frequency than the *DM-1* locus.

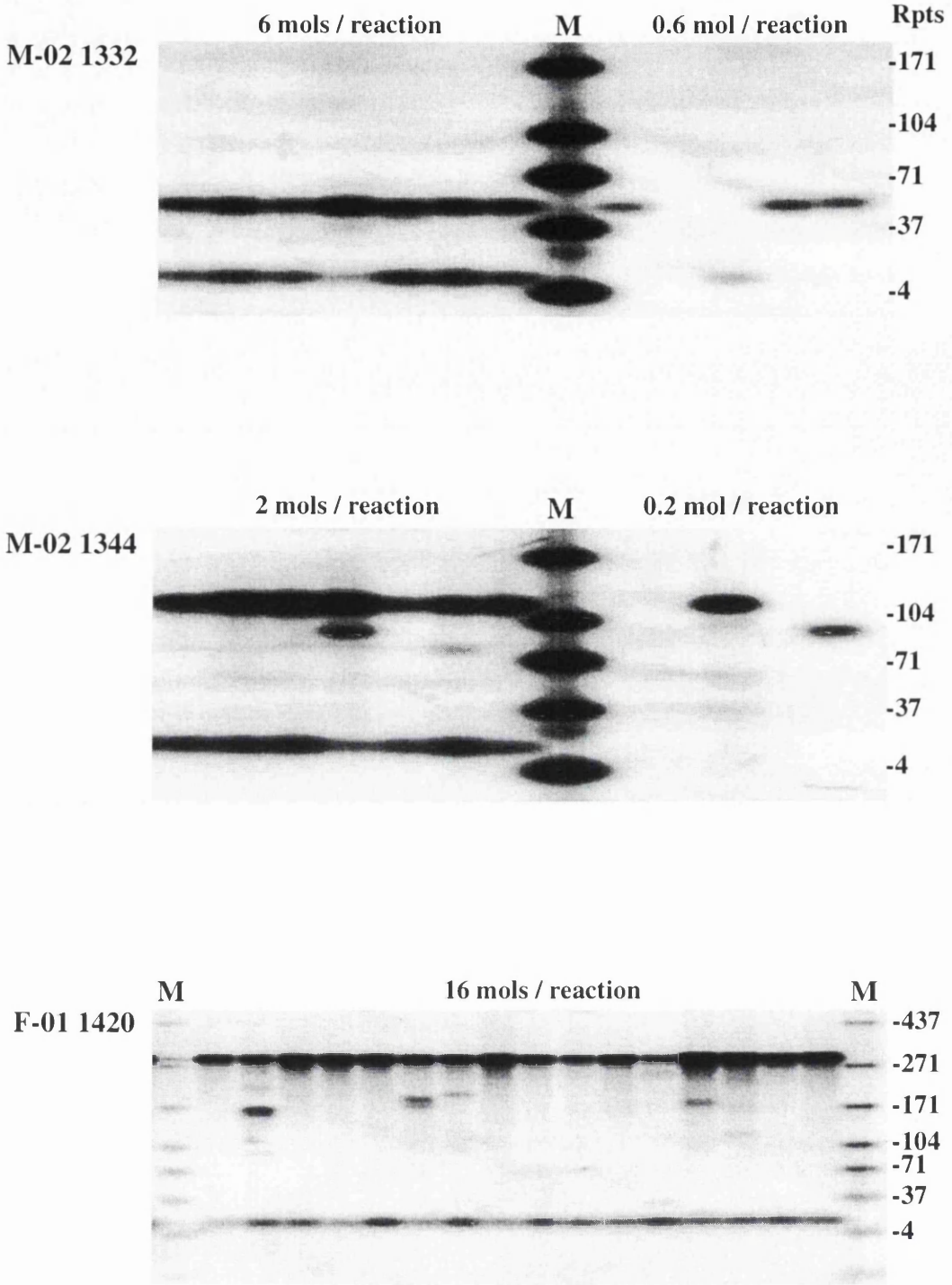

Figure 5.2 SP-PCR at the *CTG18.1* locus on lymphoblastoid cell line (CEPH) DNA. Preliminary SP-PCR at the *CTG18.1* locus on CEPH DNA samples M-02 1332 (11,52), CEPH M-02 1344 (17,118) and CEPH F-011420 (15,330). Amplisize/2.5Kb marker bands (M) are labelled as equivalent sizes in repeats. Mols = molecules. M-02 1332 has a mutation rate of 2%, M-02 1344 has a mutation rate of 8.7% and F-01 1420 has a mutation rate of 11%. Therefore there is an increasing mutation rate with increasing repeat length, biased toward contractions.

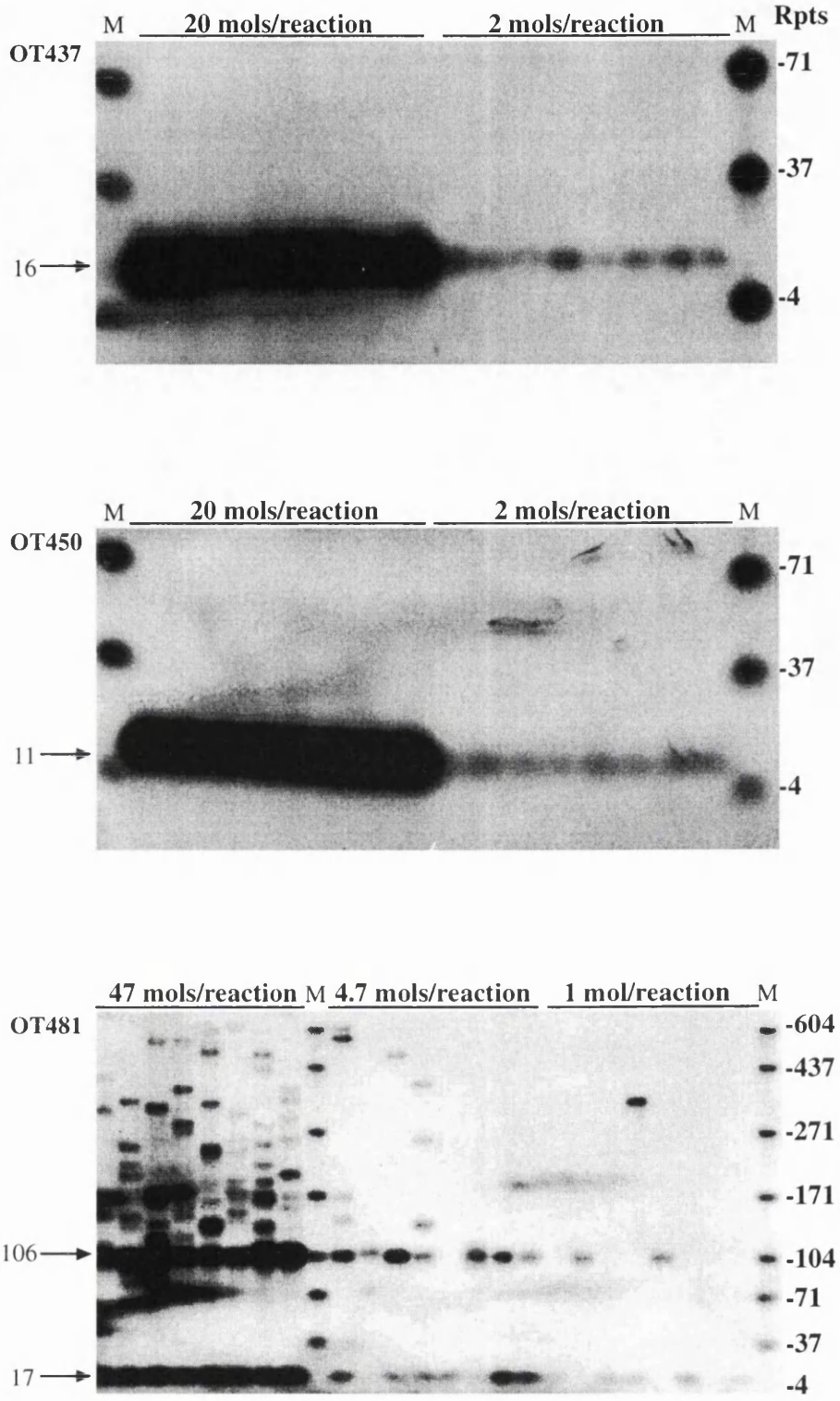

Figure 5.3 Preliminary SP-PCR at the *CTG18.1* locus on ovarian tumour DNA. Ovarian tumour samples OT437 (16,16), OT450 (11,11) and OT481 (17,106) underwent preliminary SP-PCR at the *CTG18.1* locus. Amplisize/2.5Kb marker (M) bands are labelled as equivalent sizes in repeats. Mols = molecules. Tumour OT481 is demonstrating instability of the 106 repeat allele, biased toward expansions of up to 500 repeats.

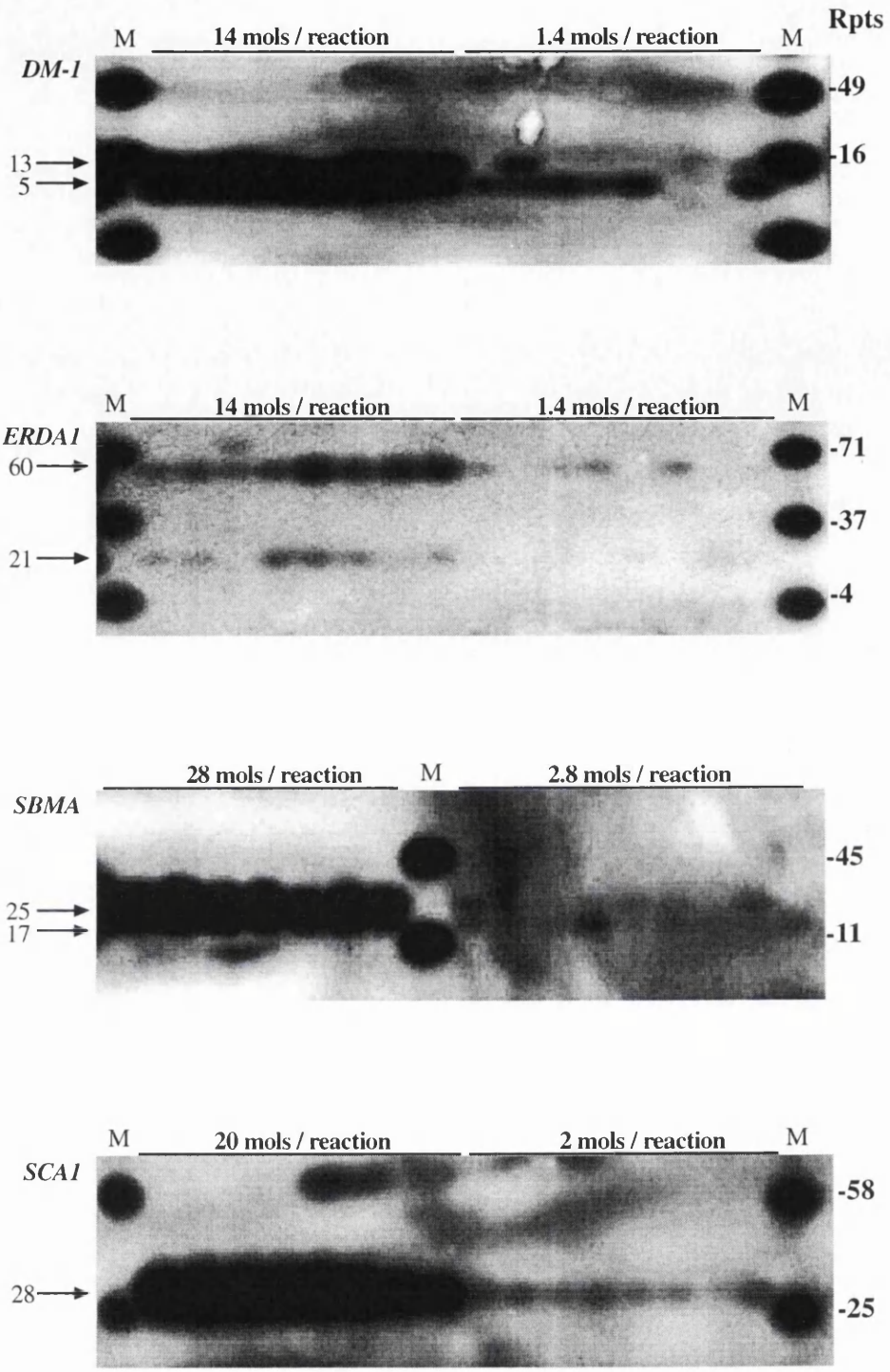

Figure 5.4 Preliminary SP-PCR at multiple trinucleotide repeat loci on ovarian tumour DNA sample OT481. Preliminary SP-PCR was carried out on ovarian tumour DNA sample OT481 at the *DM-1, ERDA 1, SBMA* and *SCA1* loci. Amplisize/2.5Kb marker (M) bands are labelled as equivalent sizes in repeats. Mols = molecules. Sizes of alleles at *DM-1 and ERDA1* were determined by fluorescent PCR with GeneScan analysis and at *SBMA* and *SCA1* by Kodak 1D Image analysis software. All loci are stable, though one expansion can be seen from the 60 repeat allele at the *ERDA1* locus.

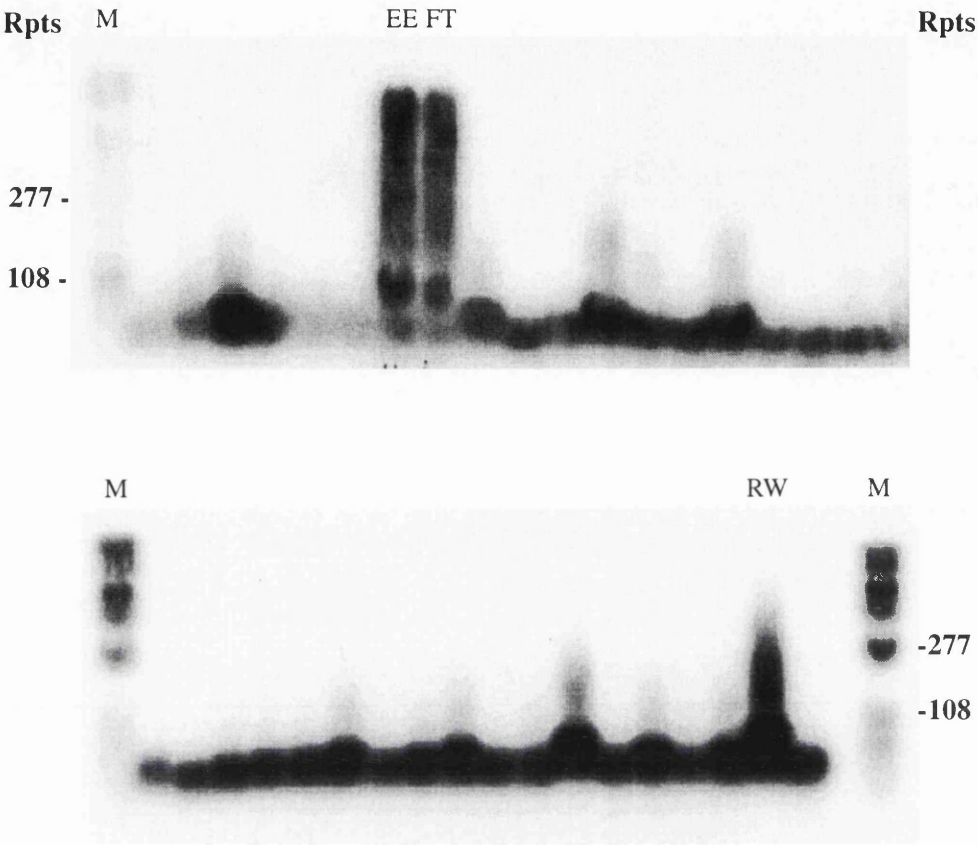

Figure 5.5 PCR Southern blot of typing PCR at the *CTG18.1* locus. PCR allele typing at the *CTG18.*1 locus was carried out on blood DNA of HNPCC individuals and the products Southern blotted. 1Kb marker (M) bands are labelled as equivalent sizes in repeats. Samples EE (16, 93) and FT (11, 93) demonstrate bands at approximately 100 repeats, but also a smear that is indicative of gross expansion. Sample RW (17, 54) also has a slight smear though SP-PCR showed this sample was stable.

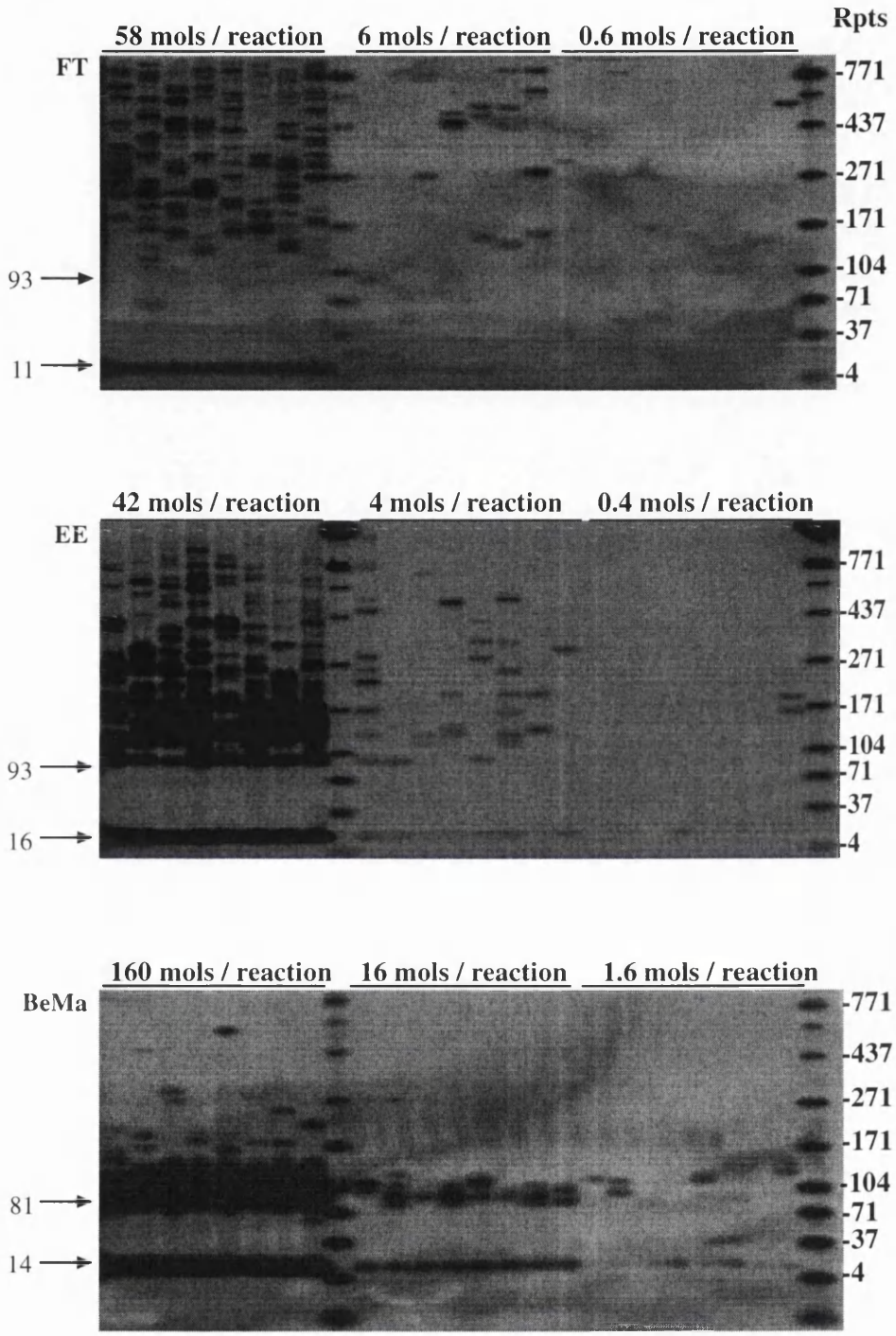

Figure 5.6 Preliminary SP-PCR at the *CTG18.1* locus on blood DNA from HNPCC individuals. Preliminary SP-PCR at the *CTG18.*1 locus on blood DNA isolated from HNPCC samples FT (11, 93), EE(16, 93) and BeMa (14, 81). Amplisize/2.5Kb marker (M) bands are labelled as equivalent sizes in repeats. Mols = molecules. The 93 repeat allele of FT and EE is very unstable and biased toward expansions. The 81 repeat allele of BeMa is more stable but with an increased mutation rate compared to the *DM*-1 blood samples analysed at the same locus in Section 5.1.

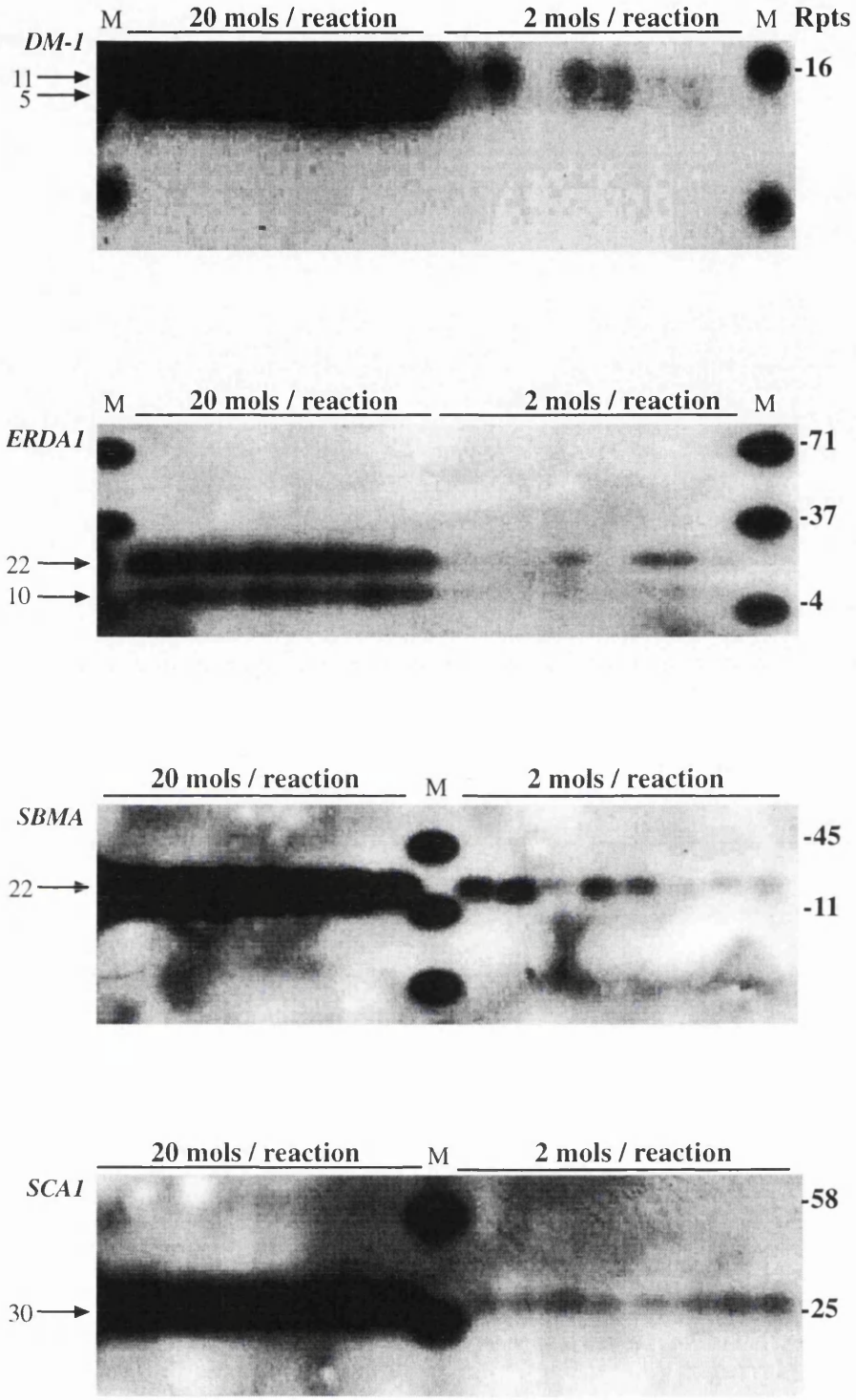

Figure 5.7 Preliminary SP-PCR at multiple trinucleotide repeat loci on HNPCC blood DNA sample FT. Preliminary SP-PCR was carried out on HNPCC blood DNA sample FT at the *DM-1, ERDA 1, SBMA* and *SCA1* loci. Amplisize/2.5Kb marker (M) bands are labelled as equivalent sizes in repeats. Mols = molecules. Sizes of alleles at *DM-1 and ERDA1* were determined by fluorescent PCR with GeneScan analysis and at *SBMA* and *SCA1* by Kodak 1D Image analysis software. All loci are stable.

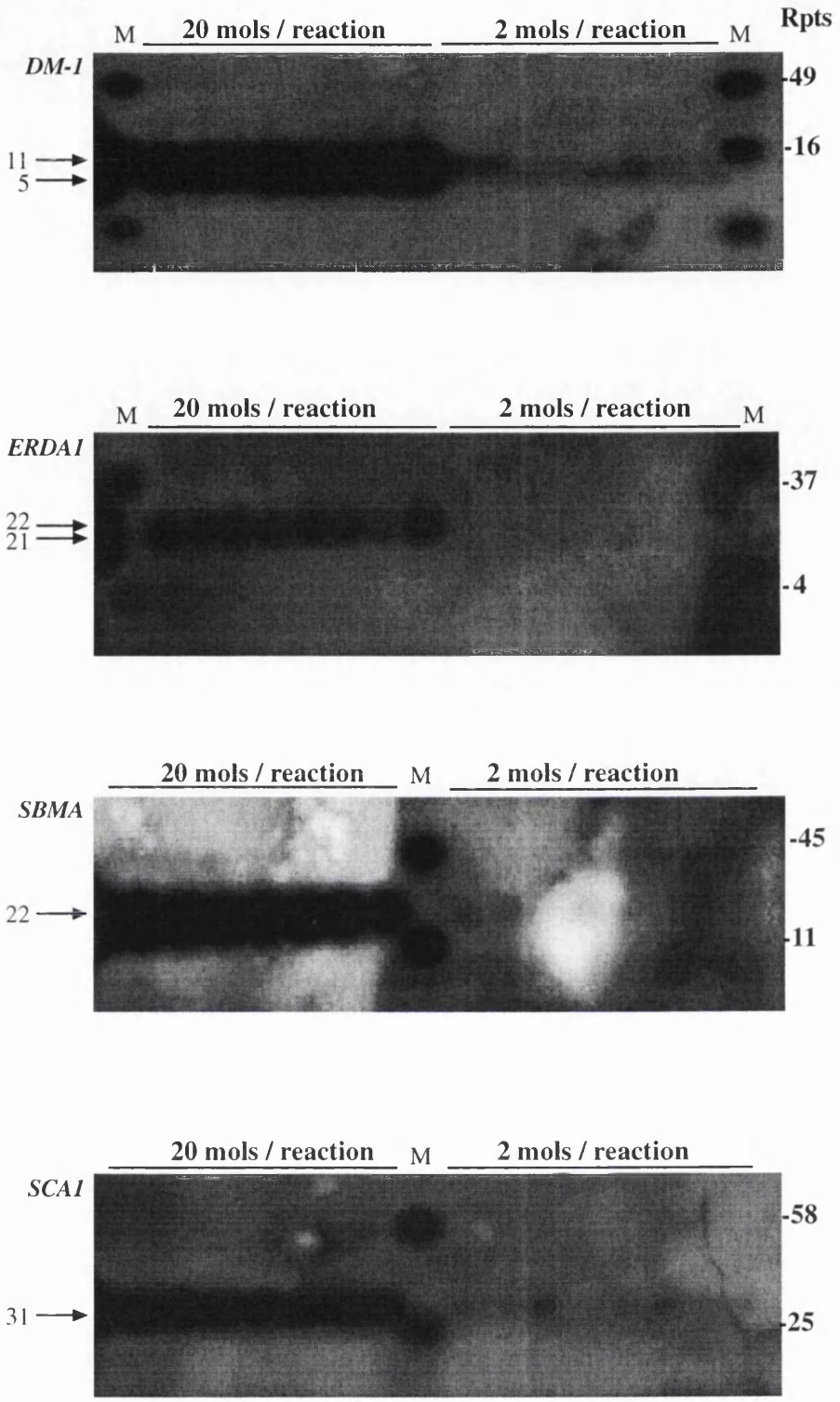

Figure 5.8 Preliminary SP-PCR at multiple trinucleotide repeat loci on HNPCC blood DNA sample EE. Preliminary SP-PCR was carried out on HNPCC blood DNA sample EE at the *DM-1, ERDA 1, SBMA* and *SCA1* loci. Amplisize/2.5Kb marker (M) bands are labelled as equivalent sizes in repeats. Mols = molecules. Sizes of alleles at *DM-1 and ERDA1* were determined by fluorescent PCR with GeneScan analysis and at *SBMA* and *SCA1* by Kodak 1D Image analysis software. All loci are stable.

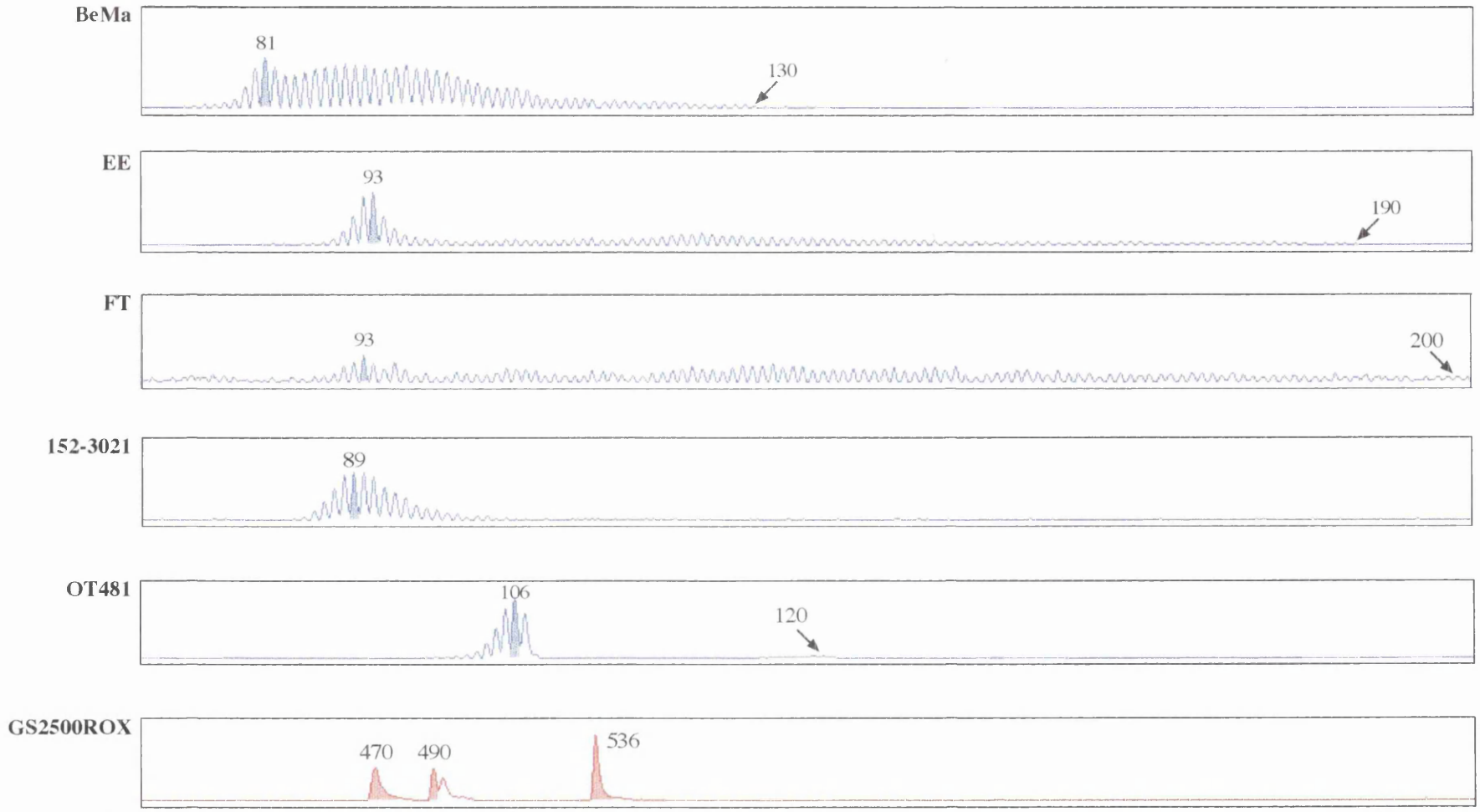

Figure 5.9 GeneScan profiles of fluorescent PCR products from expanded alleles at the *CTG18.1* locus. First round fluorescent PCR was carried out at the *CTG18.1* locus on HNPCC blood samples BeMa (14, 81), EE (16, 93) and FT (11, 93), DM-1 blood sample 152-3021 (27, 89) and ovarian tumour sample OT481 (17, 106) and the products anlaysed with GeneScan software. Displayed are the profiles of the upper allele for each sample. PCR products are blue, GS2500ROX standard is red and the peaks depicted in bp. Allele peaks are labelled as repeats. The profiles reflect what can be seen with SP-PCR. Samples 152-3021 and OT481 have tighter distributions as the majority of alleles have length sizes close to the progenitor. Samples EE and FT however have long distributions which in FT has caused the progenitor peak to be only just detectable above the other alleles.

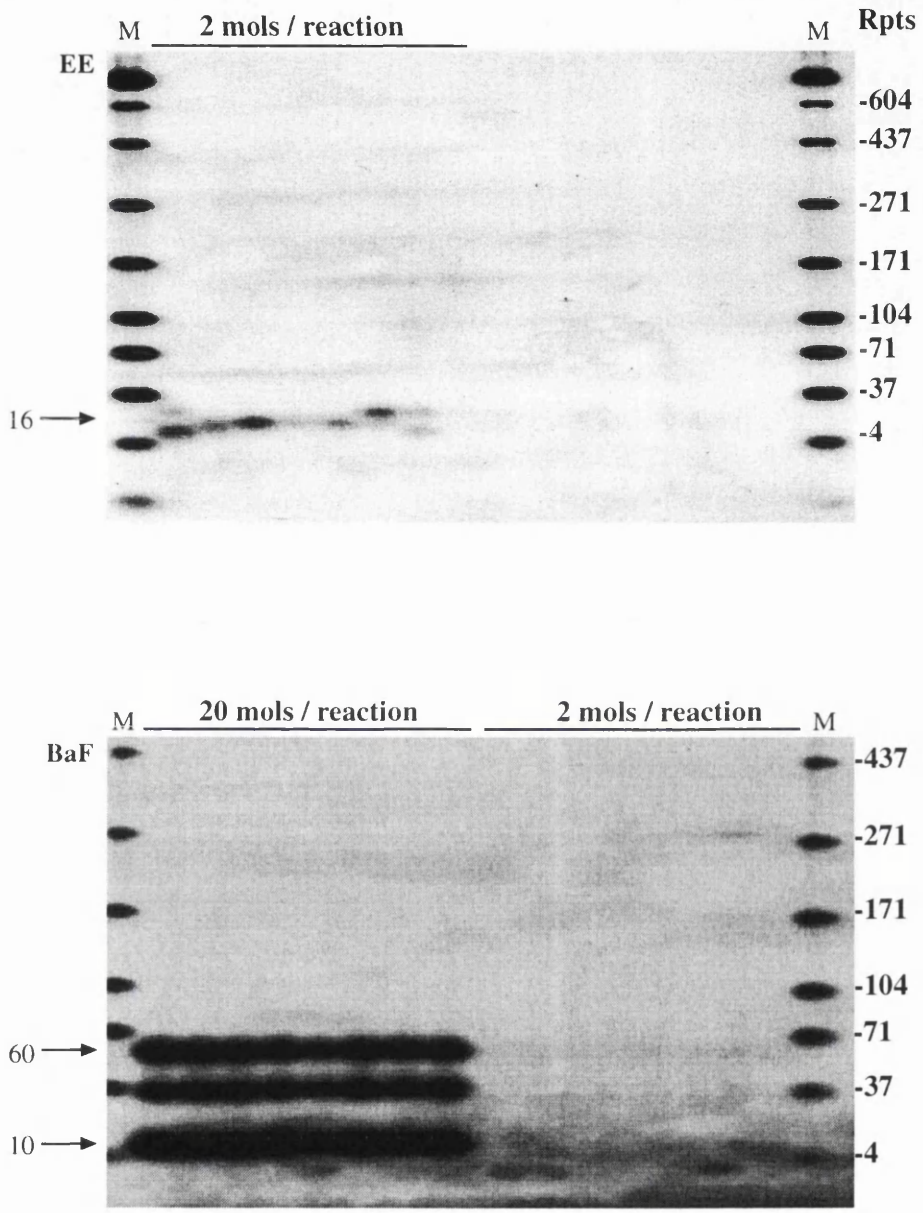

Figure 5.10 Preliminary SP-PCR on HNPCC tumour DNA samples EE and BaF. Preliminary SP-PCR was carried out on HNPCC tumour DNA sample EE (16, 93) at the *CTG18.1* locus and BaF (10, 60) at the *ERDA1* locus. Amplisize/2.5Kb marker (M) bands are labelled as equivalent sizes in repeats. Mols = molecules. The lower allele of EE appears unstable but the upper allele could not be detected. Both alleles of BaF seem stable at the *ERDA1* locus, although artefacts can be observed between the two main PCR products.

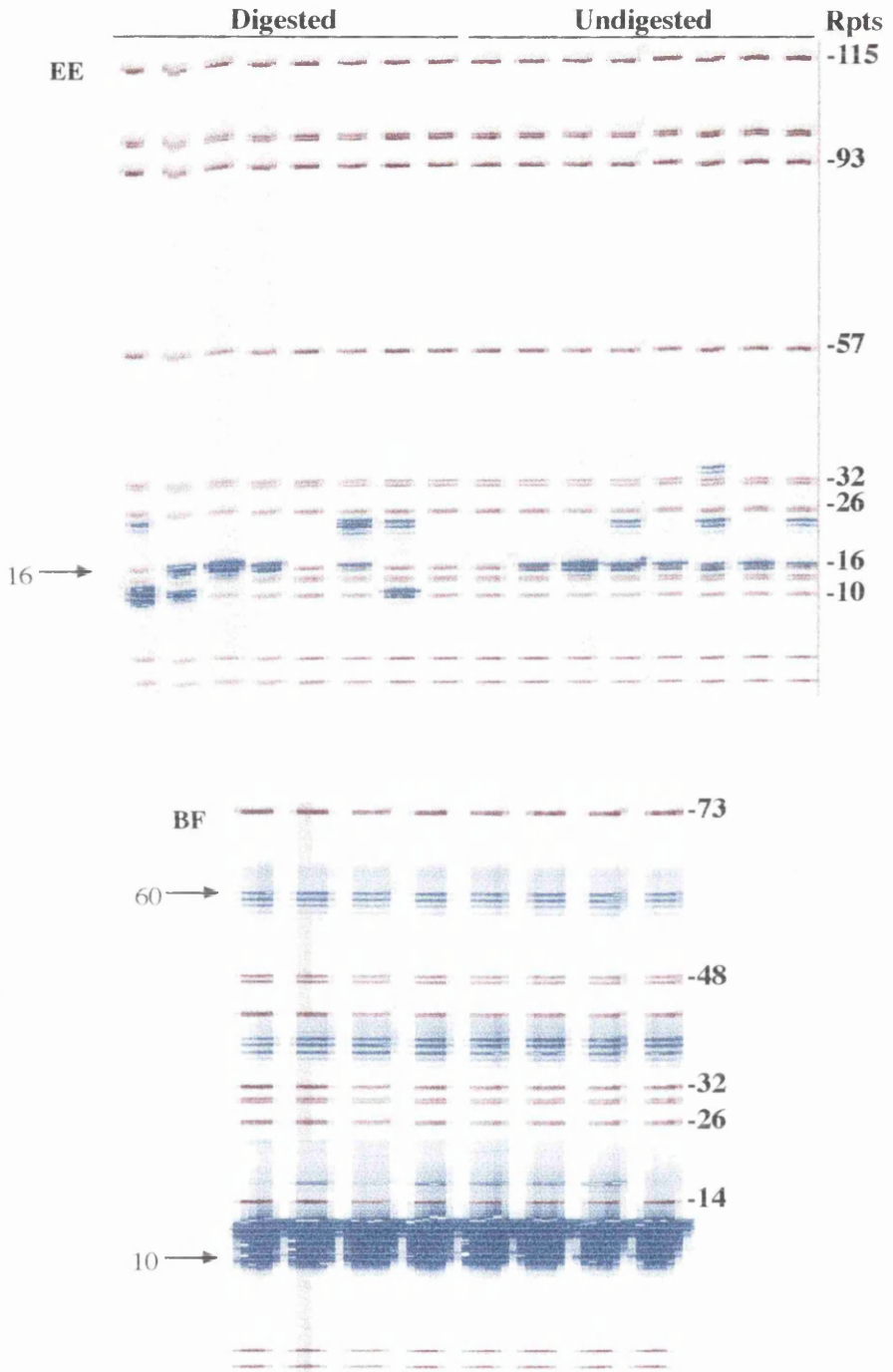

Figure 5.11 Second round fluorescent PCR on HNPCC tumour DNA. Second round fluorescent PCR at the *CTG18.1* locus on HNPCC tumour DNA sample EE (16, 93) and at the *ERDA1* locus on tumour DNA sample BF (10, 60). PCR products are blue, GS2500ROX standard is red. GS2500ROX standard marker bands are labelled as equivalent in repeats. The lower allele of EE is unstable at the *CTG18.1* locus in the tumour but the upper allele could not be detected. Both alleles of BF are stable at the *ERDA1* locus although the SP-PCR artefact is still evident.

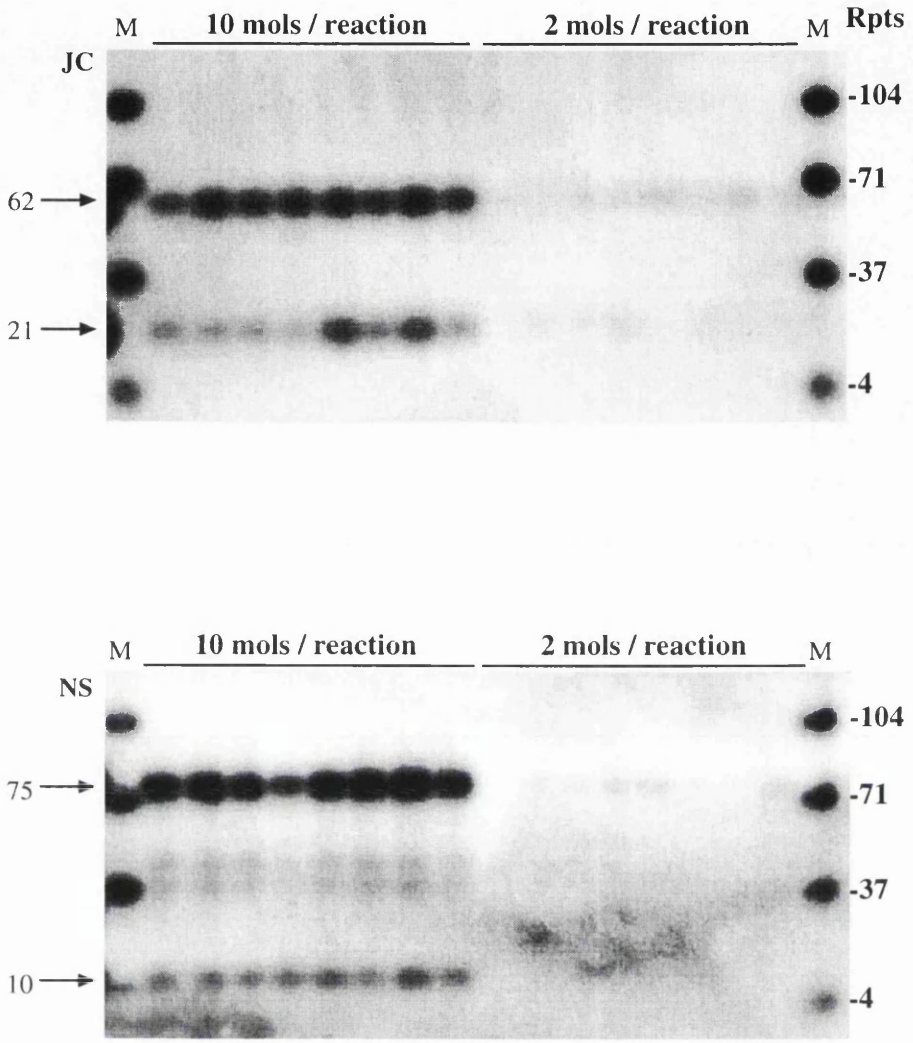

Figure 5.12 Preliminary SP-PCR at the *ERDA1* locus on blood DNA from HNPCC individuals. Preliminary SP-PCR at the *ERDA1* locus on HNPCC blood DNA samples JC (21 ,62) and NS (10, 75). Amplisize/2.5Kb marker (M) bands are labelled as equivalent sizes in repeats. Mol = molecules.Both samples are stable, although artefact products can be seen in sample NS.

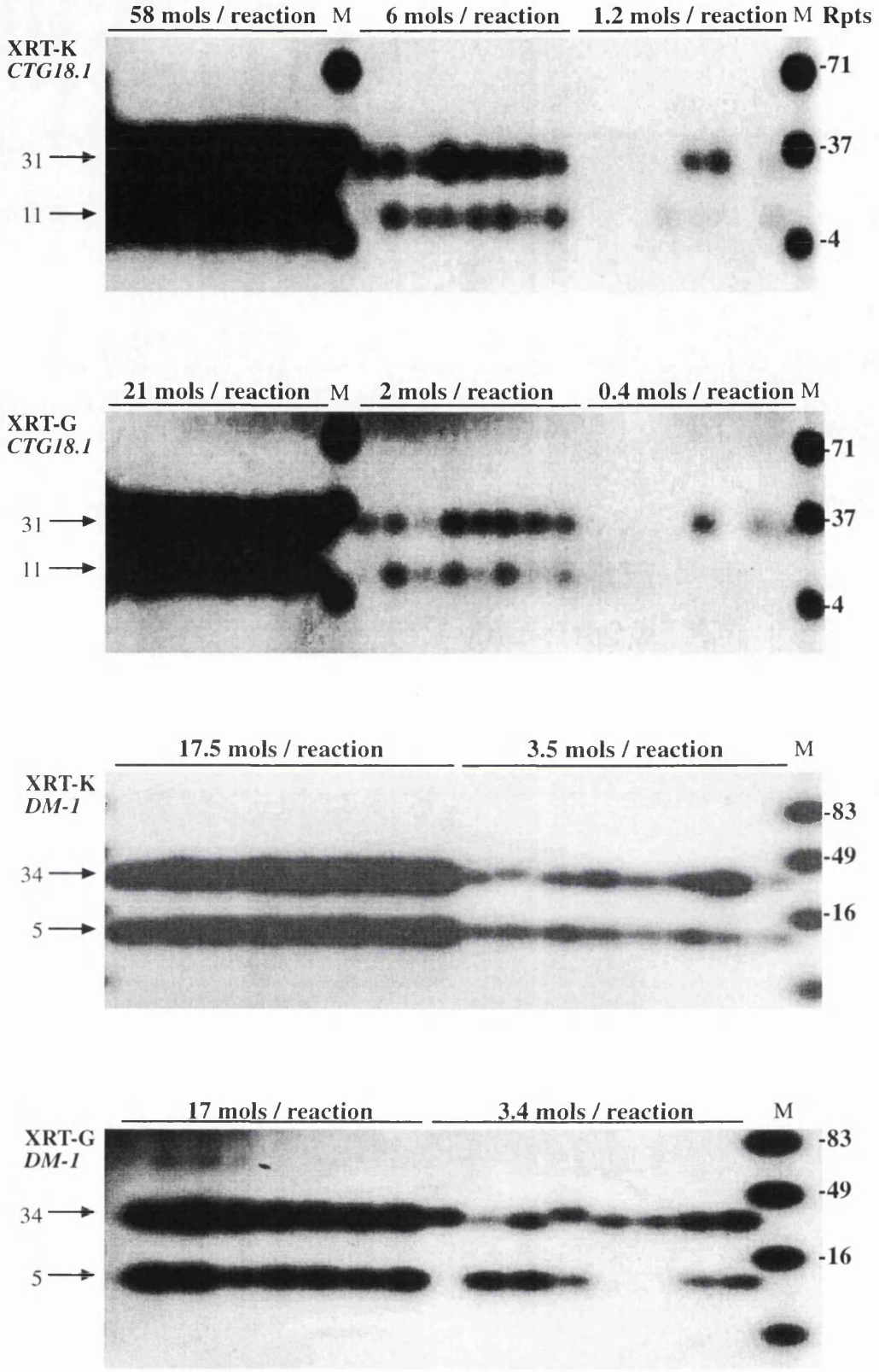

Figure 5.14 Preliminary SP-PCR at the *CTG18.1* and *DM-1* loci on sperm DNA extracted before and after radiotherapy treatment. Preliminary SP-PCR was carried out on sperm DNA extracted before (XRT-K) and after (XRT-G) radiotherapy treatment on individual 1970 (11, 31) *CTG18.1* (5, 34) *DM-1*. Amplisize/2.5Kb marker (M) bands are labelled as equivalent sizes in repeats. Mols = molecules. There are no obvious, gross size differences after radiotherapy treatment.

Sample	Locus	Progenitor Allele Length	No. of reactions	m	Total mols analysed	No. mutants	mr	A.l.c.
XRT-K 1970	*CTG18.1*	11	204/240	1.9	455	4	0.88	0.049
		31	195/240	1.67	402	23	5.7	-0.102
XRT-G 1970	*CTG18.1*	11	148/239	0.97	231	1	0.4	-0.009
		31	137/239	0.85	203	12	5.9	-0.078
XRT-K 1970	*DM-1*	5	81/90	2.3	207	0	0	0
		35	68/90	1.41	127	74	58.3	0.725
XRT-G 1970	*DM-1*	5	79/96	1.73	166	0	0	0
		35	78/96	1.67	161	84	52.2	0.348

Table 5.1 Single molecule analysis of sperm before and after radiotherapy treatment at the *CTG18.1* and *DM-1* loci. Sample XRT-1970 was typed to have (11, 31) *CTG18.1* and (5, 35) *DM-1*. XRT-K denotes before treatment, XRT-G denotes after treatment. Following first round SP-PCR, second round fluorescent PCR was carried out on the single molecule products. Positive reactions were scored according to Poisson analysis and the mutation rates and average length changes calculated.

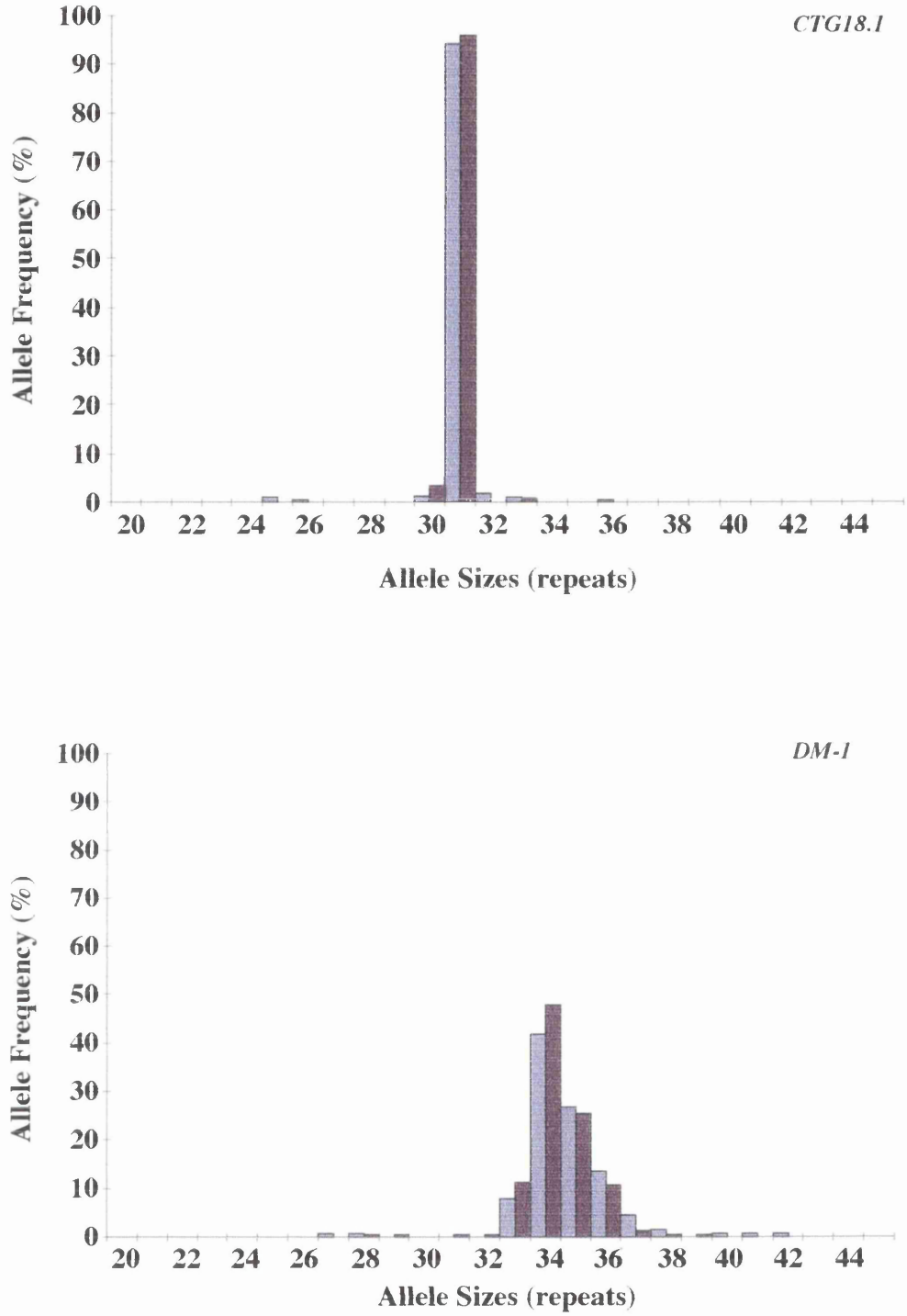

Figure 5.15 Allele distributions of the larger allele in sperm DNA from individual XRT1970 at the *CTG18.1* and *DM-1* loci before and after radiotherapy treatment. Second round fluorescent PCR products were analysed with GeneScan software and the frequencies of the detected alleles plotted above. ▪ denotes before and ▪ denotes after radiotherapy treatment. There is a slight change in progenitor allele frequencies at the two loci, but probably due to stochastic variation in the samples.

Chapter 6

Discussion

The initial aim of this project was to establish the natural behaviour of two trinucleotide repeat loci, both somatically and in the male germline, and then to observe how this behaviour could be adversely affected, if at all, by *trans*-acting factors such as mutations in mismatch repair genes, that would normally correct the repeat should any mutations occur. This was possible after the discovery of two expandable trinucleotide repeat loci known as *CTG18.1* and *ERDA1* on chromosomes 18q21.1 (Breschel *et al.*, 1997) and 17q21.3 (Nakamoto *et al.*, 1997) respectively. Locus *CTG18.1* is an expandable trinucleotide repeat in the putative third intron of the *SEF2-1* gene and was first identified in 1997 while searching for a candidate for Bipolar Affective Disorder (BPAD), although it was discounted as involved in the disorder. *CTG18.1* demonstrated a heterozygosity of 84% with the capacity to achieve allele lengths of greater than 2,000 repeats. *ERDA1* was detected by the RED method and alleles up to 115 repeats have been detected (Vincent *et al.*, 1999). Other trinucleotide repeat loci that cause disease have expanded allele frequencies and hence incidence rates of 1 in 8,000 (*DM-1*) to 1 in 10,000 (*HD*). The *CTG18.1* and *ERDA1* loci have an increased frequency of expanded alleles (>50 repeats) of approximately 3% and 15% respectively. It was therefore expected that a significant number of expanded alleles would be found when screening a large set of samples. The bank of semen samples collected from the Western Infirmary allowed PCR typing protocols to be developed and exact allele frequencies to be determined for each locus. 628 and 616 chromosomes from the semen samples were typed at the *CTG18.1* and *ERDA1* loci respectively in order to select candidates for SP-PCR analysis which would encompass both normal (<50 repeats) and expanded alleles. Initially fluorescent nucleotides were used for the PCR reactions to be analysed by GeneScan 3.1, but these were replaced with fluorescent primers due to weak signals and obscuring of the product bands by unincorporated fluorescence. Alleles ranging from 10-91 repeats and 7-82 repeats were observed at the *CTG18.1* and *ERDA1* loci respectively. Samples where only one allele was observed were checked to confirm homozygosity by PCR Southern blot, but no further alleles were detected. It is possible that very large alleles that may be present in the samples cannot be detected by PCR and so the true frequencies of the alleles could be altered slightly. This could be resolved by restriction enzyme digestion of the two loci followed by Southern blot analysis, thereby resolving any hugely expanded repeats.

The observed allele ranges of 10-91 repeats at *CTG18.1* and 7-82 repeats at *ERDA1* in this study matched the data published by Breschel *et al.* (1997), Ikeuchi *et al.* (1998), Nakamoto *et al.* (1997) and Deka *et al.* (1999) indicating that the PCR products were accurate without requiring sequencing of the repeats to confirm allele lengths. Similarly to the *DM-1* locus, the most common allele is at approximately 30-35% frequency, although this is a 5 repeat at *DM-1*, a 11 repeat at *CTG18.1* and a 10 repeat at *ERDA1*. The distribution of the alleles at *ERDA1* indicated this locus to be more stable than the *CTG18.1* locus, which in turn is more stable than the *DM-1* locus (Appendix 2). It has been well documented that alleles greater than 50 repeats rapidly expand within one generation at the *DM-1* locus, whereas similarly sized alleles are much more stable at *CTG18.1* (Breschel *et al.*, 1997) and *ERDA1* (Nakamoto *et al.*, 1997, Ikeuchi *et al.*, 1998, Deka *et al.*, 1999). This is reflected in the allele frequencies of repeats in this range, i.e. frequencies of alleles >50 repeats are 1 in 8000 at *DM-1*, 3% at *CTG18.1* and 15% at *ERDA1*. There is also a lack of alleles between approximately 40 and 50 repeats which suggests that the larger alleles arose from rare large length changes from the smaller, normal alleles.

To confirm the observations from paternal transmissions at *ERDA1* of a bias toward contractions and providing data for the *CTG18.1* locus, semen samples were selected primarily to analyse the range of 50-100 repeats, and this was extended to encompass normal sized alleles also. Large length changes could be assessed by analysing large quantities of DNA per SP-PCR reaction and expansions of 11 and 15 repeats from 35 and 38 repeat alleles at the *CTG18.1* locus, and a 25 repeat expansion from a 29 repeat allele at the *ERDA1* locus confirms that large increases in repeat length could give rise to novel expanded alleles in the population. To analyse small length changes, SP-PCR at the single molecule level underwent a second round of fluorescent PCR, then analysed by GeneScan 3.1 software. The advantage of combining two PCR techniques that allows separation of effectively single molecules and then precise sizing of the repeats enables an exact measurement to be taken of the numbers of DNA molecules (sperm) with alleles different to the progenitor, and hence mutation rates and average length changes could be calculated. In order to assess approximately the same number of molecules for each allele, a Poisson analysis was applied to the numbers of observed negative reactions in multiple PCR reactions and hence the amount of input DNA could be calculated. Blood DNA was not available from the semen donors and hence the progenitor allele in each case was determined by carrying out fluorescent PCR on the somatic DNA fraction and analysing the products with GeneScan 3.1. The single molecule data confirmed that in accordance with other trinucleotide repeats, there is an increasing mutation rate with increasing allele length at both loci. Single molecule observations for *DM-1* (Appendix 2), *HD* (Leeflang *et al.*, 1995, Leeflang *et al.*, 1999), *SCA1* (Koefoed *et al.*, 1998), *DRPLA* (Takiyama *et al.*, 1999), and *MJD/SCA3* (Takiyama *et al.*, 1997) in the male germline

suggest that *CTG18.1* and *ERDA1* are much more stable and they have a mutation pattern similar to *MJD/SCA3*. *MJD/SCA3* demonstrates an increasing frequency of contractions with increasing allele length, as do both *CTG18.1* and *ERDA1*, although *CTG18.1* seems slightly more varied. These differences may become less significant if a greater number of samples were analysed, but if not, they could be attributed to various haplotypes in the flanking sequence or variants in other genes that could be having an effect. Utilising PCR primers that specifically amplify *ERDA1* repeats with either the A or C polymorphism confirmed that alleles <50 repeats possessed the A variant whereas the larger repeats possessed the C variant. This suggests that if changes in the polymorphism predispose a repeat to increase in size, it was not detected here. This therefore suggests that the bimodal distribution of small and large alleles at the *ERDA1* locus reflects two sets of independent alleles derived from a small allele that possibly mutated immediately 3' to the repeat and subsequently underwent a large expansion. This is strengthened further by the typing of the A/C polymorphism in chimpanzee DNA which found small repeats, all associated with the A variant. Although further typing would be needed of Gorilla and Orang-utan DNA to confirm that this is the ancestral haplotype, it seems likely that the *ERDA1* alleles are derived from a small repeat with the A allele, one of which underwent a mutation immediately 3' to the repeat to the C variant, which could have predisposed it to expansion and hence establish the pool of large alleles.

It is also interesting that there is a sudden increase in the mutation rate between 35 and 40 repeats at the *CTG18.1* locus that is similar to that seen at the *DM-1* (Appendix 2) and *HD* (Leeflang *et al.*, 1995, Chong *et al.*, 1997, Leeflang *et al.*, 1999) loci. Sample GS130 (31, 40) shows a huge difference in mutation rate of 4.2% to 23.7% between the two alleles although there are only 9 repeats between them and subject to the same *trans*-acting factors within the genome of that individual. Sample GS148 (28, 35) has mutation rates of 2.8 and 7.5% respectively. Therefore, similarly to the *DM-1* locus where there is a large increase to mutation rates of 40 and 60% for 34 and 35 repeats from a 17% mutation rate at a 32 repeat allele, and the HD locus where 30, 36 and 38 repeats have mutation rates of 12, 53 and 92% (Leeflang *et al.*, 1995), the mid 30 repeat range seems to stimulate a dramatic increase in instability. It seems therefore that irrelevant of the repeat sequence, lengths of 30-40 repeats are a major factor in determining an increased mutation rate, suggesting that an increased number of secondary structures such as hairpins could form as noted with the *DM-1* repeat (Pearson and Sinden, 1998). 50% of 35 repeat DNA was found to form slipped stranded DNA (S-DNA), whereby alternative structures are induced by the repeats. Alternatively, given the fact that a 35 repeat trinucleotide has the same length as an average Okazaki fragment, slippage could be occurring in replication, leading to alternative structures in the DNA. The larger the repeat, the greater the likelihood that initiation of Okazaki fragment synthesis occurs within the repeat tract.

The mechanism of meiotic drive has been hypothesised to contribute toward unequal numbers of affected offspring in *DRPLA, MJD/SCA3* (Ikeuchi *et al.*, 1996) and disputed in *DM-1* (Carey *et al.*, 1994, Gennarelli *et al.*, 1994, Hurst *et al.*, 1995, Leeflang *et al.*, 1996). Comparing numbers of positive PCR reactions between upper and lower alleles in each individual at both loci using a Chi-squared test at a 95% confidence interval found no significant difference in the numbers of sperm with either allele. Therefore in the male germline, there is no evidence for meiotic drive in favour of either allele at each locus and hence segregation distortion could be the result of meiotic drive in females. However, in testing meiotic drive in the male germline, we are asking if there is a bias in the number of sperm with expanded trinucleotides compared to normal alleles. This does not reveal whether an expanded allele would confer some advantage on the sperm in fertilising an egg, i.e. motility and overall ability to survive for a longer period.

The behaviour of normal and expanded repeats at the *CTG18.1* and *ERDA1* loci have now been established in the male germline. However, in order to assess how these repeats could change in response to defective mismatch repair, it was important to establish the level of mutations observed in somatic tissue. Therefore, blood DNA available from *DM-1* patients were typed at *CTG18.1* and *ERDA1* and preliminary SP-PCR carried out at the *CTG18.1* locus on expanded alleles. As expected from SP-PCR at the *DM-1* locus on blood DNA (Monckton *et al.*, 1995, Martorell *et al.*, 1998), the male germline is more unstable than the blood. Alleles up to 69 repeats at *CTG18.1* were very stable and an 89 repeat allele is mutating but at a lower rate than would be expected at the *DM-1* locus. However, similarly to the *DM-1* locus, these mutations are biased toward expansions, thereby forming a definite lower boundary at the progenitor allele. The opposite could be said for the CEPH lymphoblastoid cell line DNA however, as SP-PCR at the *CTG18.1* locus demonstrated how very large expansions up to 330 repeats have a definite upper boundary and display contractions in repeat length as far as the normal range. Other publications of expanded repeat behaviour in EBV transformed lymphoblastoid cell lines also show the same behaviour (Ashizawa *et al.*, 1996), suggesting this is a feature of how repeat sequences behave in transformed cell lines. The difference between the *DM-1* and *CTG18.1* repeats is that there are some expansions seen in the cell lines from the *DM-1* repeat, but none detected in this example at the *CTG18.1* repeat. Therefore it seems as they are identical in the repeat sequence, the flanking DNA would be contributing to the type of instability observed. Observations in primary cell lines derived from the dura mater of a DM foetus found an increase in the size of the *DM-1* repeat over time (Wohrle *et al.*, 1995) and similarly in a transgene of approximately 160 *DM-1* repeats in primary cell lines derived from transgenic mouse tissues (Pereira and Monckton, unpublished).

Various cell line, blood and ovarian tumour DNA was available initially to investigate the possibility of microsatellite instability at the two loci under investigation. The parental cell line, A2780 was proficient in the MMR genes and *hMLH1* deficient derivatives were established by continuous exposure to cisplatin and selection for resistance (Anthoney *et al.*, 1996). Typing at both loci and *DM-1* uncovered instability of a 27 repeat at the *CTG18.1* locus from A2780 to the cell lines deficient in *hMLH1* and *hPMS2*. Comparison of blood and ovarian tumour DNA did not uncover any differences due to inability to resolve the repeats in the tumour sample. However, preliminary SP-PCR of ovarian tumour OT481 discovered gross instability of a 106 repeat allele at the *CTG18.1* locus with a bias toward expansions. It would have been very useful to analyse the blood DNA in this individual to help understand the behaviour of a 106 repeat in normal tissue, but unfortunately it was not available. It is assumed therefore a deficiency in the mismatch repair system is producing a RER+ phenotype detected as microsatellite instability.

To understand how expanded alleles at *CTG18.1* and *ERDA1* would behave in tissues where we know the exact status of mismatch repair, blood DNA was first obtained from HNPCC individuals that had been investigated for mutations in *hMLH1* or *hMSH2*. These samples were further supplemented by a second set of blood samples from HNPCC individuals who were also confirmed to have MMR mutations, but the precise details are not available. PCR allele typing at *CTG18.1* and *ERDA1* uncovered two blood samples from individuals who were first degree relatives with a *CTG18.1* expanded allele. A PCR Southern blot indicated that this large repeat was highly unstable which was subsequently confirmed with SP-PCR. This result was surprising as it was expected that the allele would be relatively stable in the normal tissue and potentially unstable in response to a complete deficiency in *hMLH1* in the tumour. Tumour DNA from one of the individuals, EE, proved inconclusive in that the upper allele did not resolve and the lower allele was unstable, but the amount of DNA was so low that it was not possible to have more than 1-2 molecules per SP-PCR reaction. It is likely that the DNA was damaged when the tumour was fixed in a paraffin block, although the upper allele may have expanded outside the limits of detection by PCR. A possible explanation for the phenomenon being observed in these individuals is that the *hMLH1* mutation is exerting a dominant negative effect on the wild type allele or on its partners in the mismatch repair complex and hence there is a vastly reduced function for this pathway. To exclude other possibilities such as inherited haplotypes in the family or that this is the normal behaviour for a 93 repeat, many more blood samples from normal individuals outside the family would have to be screened for *CTG18.1* expanded alleles to compare. Given that a 106 repeat in OT481 had a reduced instability compared to EE and FT for approximately an equivalent number of molecules per reaction, it is unlikely that this instability is the natural behaviour of a 93 repeat allele. It would also be interesting to investigate if other genes could be affected by the large

expansions at *CTG18.1* observed in OT481, EE and FT. The prime candidate would be *SEF2-1*, but also genes known to be involved in tumour etiology such as *BCL-2* at 18q21.3, which is close to CTG18.1.

Expanded alleles at the *ERDA1* locus were also investigated in the blood and tumour DNA of HNPCC individuals, but alleles up to 75 repeats in blood and 60 repeats in tumour were stable. Again, this demonstrates how stable this locus is although it has the same sequence of repeat as *DM-1, CTG18.1, DRPLA, HD, MJD/SCA3, SBMA*, and the spinocerebellar ataxias.

Finally, sperm DNA samples were available to answer the question of whether microsatellites could be sensitive to the effects of ionising radiation. The largest alleles detected at *CTG18.1* and *DM-1* were from sample XRT-1970. Single sperm analysis of these repeats found no difference 65 days after treatment compared to before, and the results are more likely to confirm the accuracy of combining SP-PCR with fluorescent PCR. The mutation rates and average length changes were essentially the same and the slight differences are probably due to stochastic variation rather than a positive effect of the radiotherapy. Therefore this provides evidence that the effects of radiotherapy, if any, are subtle and should not dissuade parents from considering reproducing after treatment.

The aim of this project was to characterise the behaviour and mutability of two trinucleotide repeat loci with the same repeat sequence, but with different flanking DNA. Comparing the mutation rates and average length changes of these loci in the male germline demonstrated a clear effect of the influence of the flanking sequence. The *ERDA1* locus has low %GC content in the flanking sequence, whereas *CTG18.1* has an increased %GC content. *ERDA1* is the most stable trinucleotide repeat loci yet characterised and *CTG18.1* is significantly more unstable, although not yet approaching that seen at *DM-1*. Suggestions are that the immediate flanking sequence of 100bp has a higher correlation of %GC content with instability than 500bp, but no obvious conserved motifs have been found between the most unstable loci (Brock *et al.*, 1999). The %GC content and the presence or absence of a CpG island could alter the conformational properties of the DNA or affect the methylation status which could in turn have an effect on DNA-binding proteins or the chromatin state of the DNA. Utilising the sensitive SP-PCR technique gives a clear insight into how the *CTG18.1* locus has responded dramatically to defects in mismatch repair, both in MMR deficient cell lines and HNPCC tissues, but also possibly in an ovarian tumour. This assay has also revealed a possible dominant negative mutation in the *hMLH1* gene which maintains its influence in two generations of a HNPCC family. In conclusion, we have successfully characterised the behaviour of two trinucleotide repeat loci and produced additional data to that arguing the significant effect of flanking DNA sequence on the stability of triplet repeats.

Bibliography

Aaltonen, L. A., Peltomaki, P., Leach, F. S., Sistonen, P., Pylkkanen, L., Mecklin, J. P., Jarvinen, H., Powell, S. M., Jen, J., Hamilton, S. R., Petersen, G.M., Kinzler, K.W., Vogelstein, B., and Delachapelle, A., (1993). Clues to the pathogenesis of familial colorectal cancer. Science *260*, 812-816.

Aaltonen, L. A., Peltomaki, P., Mecklin, J. P., Jarvinen, H., Jass, J. R., Green, J. S., Lynch, H. T., Watson, P., Tallqvist, G., Juhola, M., Kinzler, K.W., Vogelstein, B., and Delachapelle, A., (1994). Replication errors in benign and malignant tumors from hereditary nonpolyposis colorectal cancer patients. Cancer Research *54*, 1645-1648.

Akiyama, Y., Tsubouchi, N., and Yuasa, Y. (1997). Frequent somatic mutations of hMSH3 with reference to microsatellite instability in hereditary nonpolyposis colorectal cancers. Biochemical and Biophysical Research Communication *236*, 248-252.

Anthoney, D. A., McIlwrath, A. J., Gallagher, W. M., Edlin, A. R., and Brown, R. (1996). Microsatellite instability, apoptosis, and loss of p53 function in drug- resistant tumor cells. Cancer Research *56*, 1374-1381.

Anvret, M., Ahlberg, G., Grandell, U., Hedberg, B., Johnson, K., and Edstrom, L. (1993). Larger expansions of the CTG repeat in muscle compared to lymphocytes from patients with myotonic dystrophy. Human Molecular Genetics *2*, 1397-1400.

Armour, J. A., Brinkworth, M. H., and Kamischke, A. (1999). Direct analysis by small-pool PCR of MS205 minisatellite mutation rates in sperm after mutagenic therapies. Mutation Research *445*, 73-80.

Arzimanoglou, II, Lallas, T., Osborne, M., Barber, H., and Gilbert, F. (1996). Microsatellite instability differences between familial and sporadic ovarian cancers. Carcinogenesis *17*, 1799-1804.

Asherson, P., Walsh, C., Williams, J., Sargeant, M., Taylor, C., Clements, A., Gill, M., Owen, M., and McGuffin, P. (1994). Imprinting and anticipation. Are they relevant to genetic studies of schizophrenia? British Journal of Psychiatry *164*, 619-624.

Ashizawa, T., Monckton, D. G., Vaishnav, S., Patel, B. J., Voskova, A., and Caskey, C. T. (1996). Instability of the expanded (CTG)n repeats in the myotonin protein kinase gene in cultured lymphoblastoid cell lines from patients with myotonic dystrophy. Genomics *36*, 47-53.

Aslanidis, C., Jansen, G., Amemiya, C., Shutler, G., Mahadevan, M., Tsilfidis, C., Chen, C., Alleman, J., Wormskamp, N. G., Vooijs, M., Buxton, J., Johnson, K., Smeets, H.J.M., Lennon, GG., Carrano, A.V., Korneluk, R.G., Wieringa, B., and Dejong, P.J. (1992). Cloning of the essential myotonic dystrophy region and mapping of the putative defect. Nature *355*, 548-551.

Austin, C. P., Holder, D. J., Ma, L., Mixson, L. A., and Caskey, C. T. (1999). Mapping of hKCa3 to chromosome 1q21 and investigation of linkage of CAG repeat polymorphism to schizophrenia. Molecular Psychiatry *4*, 261-266.

Bain, G., Gruenwald, S., and Murre, C. (1993). E2A and E2-2 are subunits of B-cell-specific E2-box DNA-binding proteins. Molecular and Cellular Biology *13*, 3522-3529.

Baron, M., Risch, N., Hamburger, R., Mandel, B., Kushner, S., Newman, M., Drumer, D., and Belmaker, R. H. (1987). Genetic linkage between X-chromosome markers and bipolar affective illness. Nature *326*, 289-292.

Bassett, A. S., and Husted, J. (1997). Anticipation or ascertainment bias in schizophrenia? Penrose's familial mental illness sample. American Journal of Human Genetics *60*, 630-637.

Bengel, D., Balling, U., Stober, G., Heils, A., Li, S. H., Ross, C. A., Jungkunz, G., Franzek, E., Beckmann, H., Riederer, P., and Lesch, K. P. (1998). Distribution of the B33 CTG repeat polymorphism in a subtype of schizophrenia. European Arch Psychiatry Clinical Neuroscience *248*, 78-81.

Benson, K. F., Horwitz, M., Wolff, J., Friend, K., Thompson, E., White, S., Richards, R. I., Raskind, W. H., and Bird, T. D. (1998). CAG repeat expansion in autosomal dominant familial spastic paraparesis: novel expansion in a subset of patients. Human Molecular Genetics *7*, 1779-1786.

Berchuck, A., Kamel, A., Whitaker, R., Kerns, B., Olt, G., Kinney, R., Soper, J. T., Dodge, R., Clarke-Pearson, D. L., Marks, P., Mckenzie, S., Yin, S., and Bast, R.C. (1990). Overexpression of HER-2/neu is associated with poor survival in advanced epithelial ovarian cancer. Cancer Research *50*, 4087-4091.

Berrettini, W. H., Ferraro, T. N., Goldin, L. R., Weeks, D. E., Detera-Wadleigh, S., Nurnberger, J. I., Jr., and Gershon, E. S. (1994). Chromosome 18 DNA markers and manic-depressive illness: evidence for a susceptibility gene. Proceedings of the National Academy of Sciences U S A *91*, 5918-5921.

Bewtra, C., Watson, P., Conway, T., Read-Hippee, C., and Lynch, H. T. (1992). Hereditary ovarian cancer: a clinicopathological study. International Journal of Gynecological Pathology *11*, 180-187.

Bhattacharyya, N. P., Skandalis, A., Ganesh, A., Groden, J., and Meuth, M. (1994). Mutator phenotypes in human colorectal carcinoma cell lines. Proceedings of the National Academy of Sciences U S A *91*, 6319-6323.

Blackwood, D. H., He, L., Morris, S. W., McLean, A., Whitton, C., Thomson, M., Walker, M. T., Woodburn, K., Sharp, C. M., Wright, A. F., Shibasaki, Y., St. Clair, D. M., Porteous, D. J., and Muir, W. J. (1996). A locus for bipolar affective disorder on chromosome 4p. Nature Genetics *12*, 427-430.

Bowen, T., Guy, C. A., Craddock, N., Cardno, A. G., Williams, N. M., Spurlock, G., Murphy, K. C., Jones, L. A., Gray, M., Sanders, R. D., McCarthy, G., Chandy, K. G., Fantino, E., Kalman, K., Gutman, G. A., Gargus, J. J., Williams, J., McGuffin, P., Owen, M. J., and O'Donovan, M. C. (1998). Further support for an association between a polymorphic CAG repeat in the hKCa3 gene and schizophrenia. Molecular Psychiatry *3*, 266-269.

Boyd, J., and Rubin, S. C. (1997). Hereditary ovarian cancer: molecular genetics and clinical implications. Gynecological Oncology *64*, 196-206.

Brais, B., Bouchard, J. P., Xie, Y. G., Rochefort, D. L., Chretien, N., Tome, F. M., Lafreniere, R. G., Rommens, J. M., Uyama, E., Nohira, O., Blumen, S., Korczyn, A. D., Heutink, P., Mathieu, J., Duranceau, A., Codere, F., Fardeau, M., Rouleau, G. A., and Korcyn, A. D. (1998). Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. Nature Genetics *18*, 164-167.

Brazeau, P., Jr., Saucier, R., and Husain, S. M. (1972). Activities of hypothalamic HL-STH, of hypophyseal STH and of mimetric STH of plasma in adult and immature rats treated with enovid. Canadian Journal of Physiological Pharmacology *50*, 967-970.

Breschel, T. S., McInnis, M. G., Margolis, R. L., Sirugo, G., Corneliussen, B., Simpson, S. G., McMahon, F. J., MacKinnon, D. F., Xu, J. F., Pleasant, N., Huo, Y., Ashworth, R. G., Grundstrom, C., Grundstrom, T., Kidd, K. K., DePaulo, J. R., and Ross, C. A. (1997). A novel, heritable, expanding CTG repeat in an intron of the SEF2-1 gene on chromosome 18q21.1. Human Molecular Genetics *6*, 1855-1863.

Brock, G. J., Anderson, N. H., and Monckton, D. G. (1999). Cis-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. Human Molecular Genetics *8*, 1061-1067.

Bronner, C. E., Baker, S. M., Morrison, P. T., Warren, G., Smith, L. G., Lescoe, M. K., Kane, M., Earabino, C., Lipford, J., Lindblom, A., Tannergard, P., Bollag, R.J., Godwin, A.R., Ward, D.C., Nordenskjold, M., Fishel, R., Kolodner, R., and Liskay, R.M. (1994). Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature *368*, 258-261.

Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J. P., Hudson, T., Sohn, R., Zemelman, B., Snell, R.G., Rundle, S.A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P.S., Shaw, D.J., and Housman, D.E. (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. Cell *69*, 385.

Brown, R., Hirst, G. L., Gallagher, W. M., McIlwrath, A. J., Margison, G. P., van der Zee, A. G., and Anthoney, D. A. (1997). hMLH1 expression and cellular responses of ovarian tumour cells to treatment with cytotoxic anticancer agents. Oncogene *15*, 45-52.

Burgess, C. E., Lindblad, K., Sidransky, E., Yuan, Q. P., Long, R. T., Breschel, T., Ross, C. A., McInnis, M., Lee, P., Ginns, E. I., Lenane, M., Kumra, S., Jacobsen, L., Rapoport, J. L., and Schalling, M. (1998). Large CAG/CTG repeats are associated with childhood-onset schizophrenia. Molecular Psychiatry *3*, 321-327.

Buxton, J., Shelbourne, P., Davies, J., Jones, C., Van Tongeren, T., Aslanidis, C., de Jong, P., Jansen, G., Anvret, M., Riley, B., Williamson, R., and Johnson, K. (1992). Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. Nature *355*, 547-548.

Byrne, J., Rasmussen, S. A., Steinhorn, S. C., Connelly, R. R., Myers, M. H., Lynch, C. F., Flannery, J., Austin, D. F., Holmes, F. F., Holmes, G. E., Strong, L. C., and Mulvihill, J. J. (1998). Genetic disease in offspring of long-term survivors of childhood and adolescent cancer. American Journal of Human Genetics *62*, 45-52.

Campuzano, V., Montermini, L., Molto, M. D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., Zara, F., Canizares, J., Koutnikova, H., Bidichandani, S.I., Gellera, C., Brice, A., Trouillas, P., Demichele, G., Filla, A., Defrutos, R., Palau, F., Patel, P.I., Didonato, S., Mandel, J.L., Cocozza, S., Koenig, M., and Pandolfo, M. (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science *271*, 1423-1427.

Cancel, G., Gourfinkel-An, I., Stevanin, G., Didierjean, O., Abbas, N., Hirsch, E., Agid, Y., and Brice, A. (1998). Somatic mosaicism of the CAG repeat expansion in spinocerebellar ataxia type 3/Machado-Joseph disease. Human Mutation *11*, 23-27.

Carey, N., Johnson, K., Nokelainen, P., Peltonen, L., Savontaus, M. L., Juvonen, V., Anvret, M., Grandell, U., Chotai, K., Robertson, E., Middletonprice, H., and Malcolm, S. (1994). Meiotic drive at the myotonic dystrophy locus? Nature Genetics *6*, 117-118.

Chan, T. L., Yuen, S. T., Chung, L. P., Ho, J. W., Kwan, K. Y., Chan, A. S., Ho, J. C., Leung, S. Y., and Wyllie, A. H. (1999). Frequent microsatellite instability and mismatch repair gene mutations in young Chinese patients with colorectal cancer. Journal of the National Cancer Institute *91*, 1221-1226.

Chandy, K. G., Fantino, E., Wittekindt, O., Kalman, K., Tong, L. L., Ho, T. H., Gutman, G. A., Crocq, M. A., Ganguli, R., Nimgaonkar, V., Morris-Rosendahl, D. J., and Gargus, J. J. (1998). Isolation of a novel potassium channel gene hSKCa3 containing a polymorphic CAG repeat: a candidate for schizophrenia and bipolar disorder? Molecular Psychiatry *3*, 32-37.

Chong, S. S., Almqvist, E., Telenius, H., LaTray, L., Nichol, K., Bourdelat-Parks, B., Goldberg, Y. P., Haddad, B. R., Richards, F., Sillence, D., Greenberg, C. R., Ives, E., Van den Engh, G., Hughes, M. R., and Hayden, M. R. (1997). Contribution of DNA sequence and CAG size to mutation frequencies of intermediate alleles for Huntington disease: evidence from single sperm analyses. Human Molecular Genetics *6*, 301-309.

Chotai, J., Engstrom, C., Ekholm, B., son Berg, M. L., Adolfsson, R., and Nylander, P. O. (1995). Anticipation in Swedish families with schizophrenia. Psychiatric Genetics *5*, 181-186.

Coon, H., Hicks, A. A., Bailey, M. E., Hoff, M., Holik, J., Harvey, R. J., Johnson, K. J., Darlison, M. G., Reimherr, F., Wender, P., and Byerley, W. (1994). Analysis of GABAA receptor subunit genes in multiplex pedigrees with manic depression. Psychiatric Genetics *4*, 185-191.

Coon, H., Hoff, M., Holik, J., Hadley, D., Fang, N., Reimherr, F., Wender, P., and Byerley, W. (1996). Analysis of chromosome 18 DNA markers in multiplex pedigrees with manic depression. Biological Psychiatry *39*, 689-696.

Cooper, D. L., Lahue, R. S., and Modrich, P. (1993). Methyl-directed mismatch repair is bidirectional. Journal of Biological Chemistry *268*, 11823-11829.

Corneliussen, B., Thornell, A., Hallberg, B., and Grundstrom, T. (1991). Helix-loop-helix transcriptional activators bind to a sequence in glucocorticoid response elements of retrovirus enhancers. Journal of Virology *65*, 6084-6093.

Cuatrecasas, M., Villanueva, A., Matias-Guiu, X., and Prat, J. (1997). K-ras mutations in mucinous ovarian tumors: a clinicopathologic and molecular study of 95 cases. Cancer *79*, 1581-1586.

Cunningham, J. M., Christensen, E. R., Tester, D. J., Kim, C. Y., Roche, P. C., Burgart, L. J., and Thibodeau, S. N. (1998). Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. Cancer Research *58*, 3455-3460.

David, G., Abbas, N., Stevanin, G., Durr, A., Yvert, G., Cancel, G., Weber, C., Imbert, G., Saudou, F., Antoniou, E., Drabkin, H., Gemmill, R., Giunti, P., Benomar, A., Wood, N., Ruberg, M., Agid, Y., Mandel, J. L., and Brice, A. (1997). Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. Nature Genetics *17*, 65-70.

De bruyn, A., Souery, D., Mendelbaum, K., Mendlewicz, J., and Van Broeckhoven, C. (1996). Linkage analysis of families with bipolar illness and chromosome 18 markers. Biological Psychiatry *39*, 679-688.

De Michele, G., Cavalcanti, F., Criscuolo, C., Pianese, L., Monticelli, A., Filla, A., and Cocozza, S. (1998). Parental gender, age at birth and expansion length influence GAA repeat intergenerational instability in the X25 gene: pedigree studies and analysis of sperm from patients with Friedreich's ataxia. Human Molecular Genetics *7*, 1901-1906.

De Rooij, K. E., De Koning Gans, P. A., Roos, R. A., Van Ommen, G. J., and Den Dunnen, J. T. (1995). Somatic expansion of the (CAG)n repeat in Huntington disease brains. Human Genetics *95*, 270-274.

Deka, R., Guangyun, S., Wiest, J., Smelser, D., Chunhua, S., Zhong, Y., and Chakraborty, R. (1999). Patterns of instability of expanded CAG repeats at the ERDA1 locus in general populations. American Journal of Human Genetics *65*, 192-198.

Detera-Wadleigh, S. D., Badner, J. A., Berrettini, W. H., Yoshikawa, T., Goldin, L. R., Turner, G., Rollins, D. Y., Moses, T., Sanders, A. R., Karkera, J. D., Esterling, L. E., Zeng, J., Ferraro, T. N., Guroff, J. J., Kazuba, D., Maxwell, M. E., Nurnberger, J. I., Jr., and Gershon, E. S. (1999). A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2. Proceedings of the National Academy of Sciences U S A *96*, 5604-5609.

Detera-Wadleigh, S. D., Badner, J. A., Goldin, L. R., Berrettini, W. H., Sanders, A. R., Rollins, D. Y., Turner, G., Moses, T., Haerian, H., Muniec, D., Nurnberger, J., and Gershon, E. S. (1996). Affected-sib-pair analyses reveal support of prior evidence for a susceptibility locus for bipolar disorder, on 21q. American Journal of Human Genetics *58*, 1279-1285.

Detera-Wadleigh, S. D., Badner, J. A., Yoshikawa, T., Sanders, A. R., Goldin, L. R., Turner, G., Rollins, D. Y., Moses, T., Guroff, J. J., Kazuba, D., Maxwell, M. E., Edenberg, H. J., Foroud, T., Lahiri, D., Nurnberger, J. I., Jr., Stine, O. C., McMahon, F., Meyers, D. A., MacKinnon, D., Simpson, S., McInnis, M., DePaulo, J. R., Rice, J., Goate, A., Gershon, E. S., and *et al.* (1997). Initial genome scan of the NIMH genetics initiative bipolar pedigrees: chromosomes 4, 7, 9, 18, 19, 20, and 21q. American Journal of Medical Genetics *74*, 254-262.

Diebold, J. (1999). Molecular genetics of ovarian carcinomas. Histological Histopathology *14*, 269-277.

Diebold, J., Baretton, G., Felchner, M., Meier, W., Dopfer, K., Schmidt, M., and Lohrs, U. (1996). bcl-2 expression, p53 accumulation, and apoptosis in ovarian carcinomas. American Journal of Clinical Pathology *105*, 341-349.

Dodson, M. K., Hartmann, L. C., Cliby, W. A., DeLacey, K. A., Keeney, G. L., Ritland, S. R., Su, J. Q., Podratz, K. C., and Jenkins, R. B. (1993). Comparison of loss of heterozygosity patterns in invasive low-grade and high-grade epithelial ovarian carcinomas. Cancer Research *53*, 4456-4460.

Dorflinger, U., Pscherer, A., Moser, M., Rummele, P., Schule, R., and Buettner, R. (1999). Activation of somatostatin receptor II expression by transcription factors MIBP1 and SEF-2 in the murine brain. Molecular and Cellular Biology *19*, 3736-3747.

Dror, V., Shamir, E., Ghanshani, S., Kimhi, R., Swartz, M., Barak, Y., Weizman, R., Avivi, L., Litmanovitch, T., Fantino, E., Kalman, K., Jones, E. G., Chandy, K. G., Gargus, J. J., Gutman, G. A., and Navon, R. (1999). hKCa3/KCNN3 potassium channel gene: association of longer CAG repeats with schizophrenia in Israeli Ashkenazi Jews, expression in human tissues and localization to chromosome 1q21. Molecular Psychiatry *4*, 254-260.

Drummond, J. T., Li, G. M., Longley, M. J., and Modrich, P. (1995). Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. Science *268*, 1909-1912.

Dubrova, Y. E., Nesterov, V. N., Krouchinsky, N. G., Ostapenko, V. A., Vergnaud, G., Giraudeau, F., Buard, J., and Jeffreys, A. J. (1997). Further evidence for elevated human minisatellite mutation rate in Belarus eight years after the Chernobyl accident. Mutation Research *381*, 267-278.

Durocher, F., Shattuck-Eidens, D., McClure, M., Labrie, F., Skolnick, M. H., Goldgar, D. E., and Simard, J. (1996). Comparison of BRCA1 polymorphisms, rare sequence variants and/or missense mutations in unaffected and breast/ovarian cancer populations. Human Molecular Genetics *5*, 835-842.

Enomoto, T., Weghorst, C. M., Inoue, M., Tanizawa, O., and Rice, J. M. (1991). K-ras activation occurs frequently in mucinous adenocarcinomas and rarely in other common epithelial tumors of the human ovary. American Journal of Pathology *139*, 777-785.

Epelbaum, J., Dournaud, P., Fodor, M., and Viollet, C. (1994). The neurobiology of somatostatin. Critical Review of Neurobiology *8*, 25-44.

Eshleman, J. R., Lang, E. Z., Bowerfind, G. K., Parsons, R., Vogelstein, B., Willson, J. K., Veigl, M. L., Sedwick, W. D., and Markowitz, S. D. (1995). Increased mutation rate at the hprt locus accompanies microsatellite instability in colon cancer. Oncogene *10*, 33-37.

Eshleman, J. R., and Markowitz, S. D. (1996). Mismatch repair defects in human carcinogenesis. Human Molecular Genetics *5*, 1489-1494.

Ewald, H., Flint, T., Degn, B., Mors, O., and Kruse, T. A. (1998). A functional variant of the serotonin transporter gene in families with bipolar affective disorder. Journal of Affected Disorders *48*, 135-144.

Fishel, R., Lescoe, M. K., Rao, M. R., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M., and Kolodner, R. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell *75*, 1027-1038.

Flynn, G. A., Hirst, M. C., Knight, S. J., Macpherson, J. N., Barber, J. C., Flannery, A. V., Davies, K. E., and Buckle, V. J. (1993). Identification of the FRAXE fragile site in two families ascertained for X linked mental retardation. Journal of Medical Genetics *30*, 97-100.

Foulkes, W. D., Black, D. M., Stamp, G. W., Solomon, E., and Trowsdale, J. (1993). Very frequent loss of heterozygosity throughout chromosome 17 in sporadic ovarian carcinoma. International Journal of Cancer *54*, 220-225.

Frank, T. S., Bartos, R. E., Haefner, H. K., Roberts, J. A., Wilson, M. D., and Hubbell, G. P. (1994). Loss of heterozygosity and overexpression of the p53 gene in ovarian carcinoma. Modern Pathology *7*, 3-8.

Freimer, N. B., Reus, V. I., Escamilla, M. A., McInnes, L. A., Spesny, M., Leon, P., Service, S. K., Smith, L. B., Silva, S., Rojas, E., Gallegos, A., Meza, L., Fournier, E., Baharloo, S., Blankenship, K., Tyler, D. J., Batki, S., Vinogradov, S., Weissenbach, J., Barondes, S. H., and Sandkuijl, L. A. (1996). Genetic mapping using haplotype, association and linkage methods suggests a locus for severe bipolar disorder (BPI) at 18q22-q23. Nature Genetics *12*, 436-441.

Fu, Y. H., Kuhl, D. P., Pizzuti, A., Pieretti, M., Sutcliffe, J. S., Richards, S., Verkerk, A. J., Holden, J. J., Fenwick, R. G., and Warren, S. T. e. a. (1991). Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell *67*, 1047-1058.

Fujita, M., Enomoto, T., Yoshino, K., Nomura, T., Buzard, G. S., Inoue, M., and Okudaira, Y. (1995). Microsatellite instability and alterations in the hMSH2 gene in human ovarian cancer. International Journal of Cancer *64*, 361-366.

Geeta Vani, R., Varghese, C. M., and Rao, M. R. (1999). Cloning of rat MLH1 and expression analysis of MSH2, MSH3, MSH6, and MLH1 during spermatogenesis. Genomics *62*, 460-467.

Gelernter, J. (1995). Genetics of bipolar affective disorder: time for another reinvention? American Journal of Human Genetics *56*, 1262-1266.
Gennarelli, M., Dallapiccola, B., Baiget, M., Martorell, L., and Novelli, G. (1994). Meiotic drive at the myotonic dystrophy locus. Journal of Medical Genetics *31*, 980.

Gouw, L. G., Castaneda, M. A., McKenna, C. K., Digre, K. B., Pulst, S. M., Perlman, S., Lee, M. S., Gomez, C., Fischbeck, K., Gagnon, D., Storey, E., Bird, T., Jeri, F. R., and Ptacek, L. J. (1998). Analysis of the dynamic mutation in the SCA7 gene shows marked parental effects on CAG repeat transmission. Human Molecular Genetics *7*, 525-532.

Grilley, M., Welsh, K. M., Su, S. S., and Modrich, P. (1989). Isolation and characterization of the Escherichia coli mutL gene product. Journal of Biological Chemistry *264*, 1000-1004.

Grimm, L., Holinski-Feder, E., Teodoridis, J., Scheffer, B., Schindelhauer, D., Meitinger, T., and Ueffing, M. (1998). Analysis of the human GDNF gene reveals an inducible promoter, three exons, a triplet repeat within the 3'-UTR and alternative splice products. Human Molecular Genetics *7*, 1873-1886.

Guy, C., Bowen, T., Daniels, J. K., Speight, G., McKeon, P., Mynett-Johnson, L., Claffey, E., McGuffin, P., Owen, M. J., Craddock, N., and O'Donovan, M. C. (1997). Exclusion of expansion of 50 CAG/CTG trinucleotide repeats in bipolar disorder. American Journal of Psychiatry *154*, 1146-1147.

Guy, C. A., Bowen, T., Jones, I., McCandless, F., Owen, M. J., Craddock, N., and O'Donovan, M. C. (1999). CTG18.1 and ERDA-1 CAG/CTG repeat size in bipolar disorder. Neurobiological Disorders *6*, 302-307.

Han, H. J., Maruyama, M., Baba, S., Park, J. G., and Nakamura, Y. (1995). Genomic structure of human mismatch repair gene, hMLH1, and its mutation analysis in patients with hereditary non-polyposis colorectal cancer (HNPCC). Human Molecular Genetics *4*, 237-242.

Harley, H. G., Rundle, S. A., Reardon, W., Myring, J., Crow, S., Brook, J. D., Harper, P. S., and Shaw, D. J. (1992). Unstable DNA sequence in myotonic dystrophy. Lancet *339*, 1125-1128.

Hawi, Z., Mynett-Johnson, L., Murphy, V., Straub, R. E., Kendler, K. S., Walsh, D., McKeon, P., and Gill, M. (1999). No evidence to support the association of the potassium channel gene hSKCa3 CAG repeat with schizophrenia or bipolar disorder in the Irish population. Molecular Psychiatry *4*, 488-491.

Henriksen, R., Wilander, E., and Oberg, K. (1995). Expression and prognostic significance of Bcl-2 in ovarian tumours. British Journal of Cancer *72*, 1324-1329.

Henthorn, P., Kiledjian, M., and Kadesch, T. (1990). Two distinct transcription factors that bind the immunoglobulin enhancer microE5/kappa 2 motif. Science *247*, 467-470.

Henthorn, P., McCarrick-Walmsley, R., and Kadesch, T. (1990). Sequence of the cDNA encoding ITF-2, a positive-acting transcription factor. Nucleic Acids Research *18*, 678.

Hermann, S., Saarikettu, J., Onions, J., Hughes, K., and Grundstrom, T. (1998). Calcium regulation of basic helix-loop-helix transcription factors. Cell Calcium *23*, 135-142.

Holmes, S. E., O'Hearn, E. E., McInnis, M. G., Gorelick-Feldman, D. A., Kleiderlein, J. J., Callahan, C., Kwak, N. G., Ingersoll-Ashworth, R. G., Sherr, M., Sumner, A. J., Sharp, A. H., Ananth, U., Seltzer, W. K., Boss, M. A., Vieria-Saecker, A. M., Epplen, J. T., Riess, O., Ross, C. A., and Margolis, R. L. (1999). Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. Nature Genetics *23*, 391-392.

Hovatta, I., Varilo, T., Suvisaari, J., Terwilliger, J. D., Ollikainen, V., Arajarvi, R., Juvonen, H., Kokko-Sahin, M. L., Vaisanen, L., Mannila, H., Lonnqvist, J., and Peltonen, L. (1999). A Genomewide Screen for Schizophrenia Genes in an Isolated Finnish Subpopulation, Suggesting Multiple Susceptibility Loci. American Journal of Human Genetics *65*, 1114-1124.

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

Hurst, G. D., Hurst, L. D., and Barrett, J. A. (1995). Meiotic drive and myotonic dystrophy. Nature Genetics *10*, 132-133.

Ikeuchi, T., Igarashi, S., Takiyama, Y., Onodera, O., Oyake, M., Takano, H., Koide, R., Tanaka, H., and Tsuji, S. (1996). Non-Mendelian transmission in dentatorubral-pallidoluysian atrophy and Machado-Joseph disease: the mutant allele is preferentially transmitted in male meiosis. American Journal of Human Genetics *58*, 730-733.

Ikeuchi, T., Sanpei, K., Takano, H., Sasaki, H., Tashiro, K., Cancel, G., Brice, A., Bird, T. D., Schellenberg, G. D., Pericak-Vance, M. A., Welsh-Bohmer, K. A., Clark, L. N., Wilhelmsen, K., and Tsuji, S. (1998). A novel long and unstable CAG/CTG trinucleotide repeat on chromosome 17q. Genomics *49*, 321-326.

Imbert, G., Saudou, F., Yvert, G., Devys, D., Trottier, Y., Garnier, J. M., Weber, C., Mandel, J. L., Cancel, G., Abbas, N., Durr, A., Didierjean, O., Stevanin, G., Agid, Y., and Brice, A. (1996). Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. Nature Genetics *14*, 285-291.

Imbert, G., Trottier, Y., Beckmann, J., and Mandel, J. L. (1994). The gene for the TATA binding protein (TBP) that contains a highly polymorphic protein coding CAG repeat maps to 6q27. Genomics *21*, 667-668.

Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D., and Perucho, M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature *363*, 558-561.

Ishikawa, K., Watanabe, M., Yoshizawa, K., Fujita, T., Iwamoto, H., Yoshizawa, T., Harada, K., Nakamagoe, K., Komatsuzaki, Y., Satoh, A., Doi, M., Ogata, T., Kanazawa, I., Shoji, S., and Mizusawa, H. (1999). Clinical, neuropathological, and molecular study in two families with spinocerebellar ataxia type 6 (SCA6). Journal of Neurological and Neurosurgical Psychiatry *67*, 86-89.

Ito, Y., Tanaka, F., Yamamoto, M., Doyu, M., Nagamatsu, M., Riku, S., Mitsuma, T., and Sobue, G. (1998). Somatic mosaicism of the expanded CAG trinucleotide repeat in mRNAs for the responsible gene of Machado-Joseph disease (MJD), dentatorubral- pallidoluysian atrophy (DRPLA), and spinal and bulbar muscular atrophy (SBMA). Neurochemical Research *23*, 25-32.

Jacobs, I. J., Smith, S. A., Wiseman, R. W., Futreal, P. A., Harrington, T., Osborne, R. J., Leech, V., Molyneux, A., Berchuck, A., Ponder, B. A., and Bast, R.C. (1993). A deletion unit on chromosome 17q in epithelial ovarian tumors distal to the familial breast/ovarian cancer locus. Cancer Research *53*, 1218-1221.

Janezic, S. A., Ziogas, A., Krumroy, L. M., Krasner, M., Plummer, S. J., Cohen, P., Gildea, M., Barker, D., Haile, R., Casey, G., and Anton-Culver, H. (1999). Germline BRCA1 alterations in a population-based series of ovarian cancer cases. Human Molecular Genetics *8*, 889-897.

Jansen, G., Willems, P., Coerwinkel, M., Nillesen, W., Smeets, H., Vits, L., Howeler, C., Brunner, H., and Wieringa, B. (1994). Gonosomal mosaicism in myotonic dystrophy patients: involvement of mitotic events in (CTG)n repeat variation and selection against extreme expansion in sperm. American Journal of Human Genetics *54*, 575-585.

Javaux, F., Donda, A., Vassart, G., and Christophe, D. (1991). Cloning and sequence analysis of TFE, a helix-loop-helix transcription factor able to recognize the thyroglobulin gene promoter in vitro. Nucleic Acids Research *19*, 1121-1127.

Jean, M., Pelletier, J., Hilpert, M., Belzile, F., and Kunze, R. (1999). Isolation and characterization of AtMLH1, a MutL homologue from Arabidopsis thaliana. Mol Gen Genet *262*, 633-642.

Jeon, H. M., Lynch, P. M., Howard, L., Ajani, J., Levin, B., and Frazier, M. L. (1996). Mutation of the hMSH2 gene in two families with hereditary nonpolyposis colorectal cancer. Human Mutation *7*, 327-333.

Jinnai, K., Sugio, T., Mitani, M., Hashimoto, K., and Takahashi, K. (1999). Elongation of (CTG)n repeats in myotonic dystrophy protein kinase gene in tumors associated with myotonic dystrophy patients. Muscle and Nerve *22*, 1271-1274.

Jones, C., Slijepcevic, P., Marsh, S., Baker, E., Langdon, W. Y., Richards, R. I., and Tunnacliffe, A. (1994). Physical linkage of the fragile site FRA11B and a Jacobsen syndrome chromosome deletion breakpoint in 11q23.3. Human Molecular Genetics *3*, 2123-2130.

Jonsson, E. G., Nimgaonkar, V. L., Zhang, X. R., Shaw, S. H., Burgert, E., Crocq, M. A., Chakravarti, A., and Sedvall, G. C. (1999). Trend for an association between schizophrenia and D3S1310, a marker in proximity to the dopamine D3 receptor gene. American Journal of Medical Genetics *88*, 352-357.

Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M., Akiguchi, I., Kimura, J., Narumiya, S., and Kakizuka, A. (1994). CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nature Genetics *8*, 221-228.

King, B. L., Carcangiu, M. L., Carter, D., Kiechle, M., Pfisterer, J., Pfleiderer, A., and Kacinski, B. M. (1995). Microsatellite instability in ovarian neoplasms. British Journal of Cancer *72*, 376-382.

Kinoshita, M., Igarashi, A., Komori, T., Tamura, H., Hayashi, M., Kinoshita, K., Deguchi, T., and Hirose, K. (1997). Differences in CTG triplet repeat expansions in an ovarian cancer and cyst from a patient with myotonic dystrophy. Muscle and Nerve *20*, 622-624.

Kinoshita, M., Takahashi, R., Hasegawa, T., Komori, T., Nagasawa, R., Hirose, K., and Tanabe, H. (1996). (CTG)n expansions in various tissues from a myotonic dystrophy patient. Muscle and Nerve *19*, 240-242.

Kirchweger, R., Zeillinger, R., Schneeberger, C., Speiser, P., Louason, G., and Theillet, C. (1994). Patterns of allele losses suggest the existence of five distinct regions of LOH on chromosome 17 in breast cancer. International Journal of Cancer *56*, 193-199.

Kirkpatrick, D. T. (1999). Roles of the DNA mismatch repair and nucleotide excision repair proteins during meiosis. Cellular and Molecular Life Sciences *55*, 437-449.

Koefoed, P., Hasholt, L., Fenger, K., Nielsen, J. E., Eiberg, H., Buschard, K., and Sorensen, S. A. (1998). Mitotic and meiotic instability of the CAG trinucleotide repeat in spinocerebellar ataxia type 1. Human Genetics *103*, 564-569.

Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hayashi, T., Saito, M., Tomoda, A., Miike, T., Naito, H., Ikuta, F., and Tsuji, S. (1994). Unstable expansion of CAG repeat in hereditary dentatorubral- pallidoluysian atrophy (DRPLA). Nature Genetics *6*, 9-13.

Koide, R., Kobayashi, S., Shimohata, T., Ikeuchi, T., Maruyama, M., Saito, M., Yamada, M., Takahashi, H., and Tsuji, S. (1999). A neurological disease caused by an expanded CAG trinucleotide repeat in the TATA-binding protein gene: a new polyglutamine disease? Human Molecular Genetics *8*, 2047-2053.

Kolodner, R. D., Hall, N. R., Lipford, J., Kane, M. F., Morrison, P. T., Finan, P. J., Burn, J., Chapman, P., Earabino, C., Merchant, E., and Bishop, D.T. (1995). Structure of the human MLH1 locus and analysis of a large hereditary nonpolyposis colorectal carcinoma kindred for mlh1 mutations. Cancer Research *55*, 242-248.

Koob, M. D., Moseley, M. L., Schut, L. J., Benzow, K. A., Bird, T. D., Day, J. W., and Ranum, L. P. (1999). An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). Nature Genetics *21*, 379-384.

Kuismanen, S. A., Holmberg, M. T., Salovaara, R., Schweizer, P., Aaltonen, L. A., de La Chapelle, A., Nystrom-Lahti, M., and Peltomaki, P. (1999). Epigenetic phenotypes distinguish microsatellite-stable and -unstable colorectal cancers. Proceedings of the National Academy of Sciences U S A *96*, 12661-12666.

Kunst, C. B., Leeflang, E. P., Iber, J. C., Arnheim, N., and Warren, S. T. (1997). The effect of FMR1 CGG repeat interruptions on mutation frequency as measured by sperm typing. Journal of Medical Genetics *34*, 627-631.

Kupryjanczyk, J., Bell, D. A., Dimeo, D., Beauchamp, R., Thor, A. D., and Yandell, D. W. (1995). p53 gene analysis of ovarian borderline tumors and stage I carcinomas. Human Pathology *26*, 387-392.

La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E., and Fischbeck, K. H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature *352*, 77-79.

Lahue, R. S., Au, K. G., and Modrich, P. (1989). DNA mismatch correction in a defined system. Science *245*, 160-164.

Lamberti, C., Kruse, R., Ruelfs, C., Caspari, R., Wang, Y., Jungck, M., Mathiak, M., Malayeri, H. R., Friedl, W., Sauerbruch, T., and Propping, P. (1999). Microsatellite instability-a useful diagnostic tool to select patients at high risk for hereditary non-polyposis colorectal cancer: a study in different groups of patients with colorectal cancer. Gut *44*, 839-843.

Laurent, C., Zander, C., Thibaut, F., Bonnet-Brilhault, F., Chavand, O., Jay, M., Samolyk, D., Petit, M., Martinez, M., Campion, D., Neri, C., Mallet, J., and Cann, H. (1998). Anticipation in schizophrenia: no evidence of expanded CAG/CTG repeat sequences in French families and sporadic cases. American Journal of Medical Genetics *81*, 342-346.

Leach, F. S., Nicolaides, N. C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomaki, P., Sistonen, P., Aaltonen, L. A., Nystrom-Lahti, M., Guan, X.Y., Zhang, J., Meltzer, P.S., Yu, J.W., Kao, F.T., Chen, D.J., Cerosaletti, K.M., Fournier, R.E.K., Todd, S., Lewis, T., Leach, R.J., Naylor, S.L., Weissenbach, J., Mecklin, J.P., Jarvinen, H., Petersen, G.M., Hamilton, S.R., Green, J., Jass, J., Watson, P., Lynch, H.T., Trent, J.M., Delachapelle, A., Kinzler, K.W., and Vogelstein, B. (1993). Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell *75*, 1215-1225.

Leeflang, E. P., McPeek, M. S., and Arnheim, N. (1996). Analysis of meiotic segregation, using single-sperm typing: meiotic drive at the myotonic dystrophy locus. American Journal of Human Genetics *59*, 896-904.

Leeflang, E. P., Tavare, S., Marjoram, P., Neal, C. O., Srinidhi, J., MacFarlane, H., MacDonald, M. E., Gusella, J. F., de Young, M., Wexler, N. S., and Arnheim, N. (1999). Analysis of germline mutation spectra at the Huntington's disease locus supports a mitotic mutation mechanism. Human Molecular Genetics *8*, 173-183.

Leeflang, E. P., Zhang, L., Tavare, S., Hubert, R., Srinidhi, J., MacDonald, M. E., Myers, R. H., de Young, M., Wexler, N. S., Gusella, J. F., and Arnheim, N. (1995). Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: quantification of the mutation frequency spectrum. Human Molecular Genetics *4*, 1519-1526.

Levinson, D. F., Mahtani, M. M., Nancarrow, D. J., Brown, D. M., Kruglyak, L., Kirby, A., Hayward, N. K., Crowe, R. R., Andreasen, N. C., Black, D. W., Silverman, J. M., Endicott, J., Sharpe, L., Mohs, R. C., Siever, L. J., Walters, M. K., Lennon, D. P., Jones, H. L., Nertney, D. A., Daly, M. J., Gladis, M., and Mowry, B. J. (1998). Genome scan of schizophrenia. American Journal of Psychiatry *155*, 741-750.

Li, T., Vallada, H. P., Liu, X., Xie, T., Tang, X., Zhao, J., O'Donovan, M. C., Murray, R. M., Sham, P. C., and Collier, D. A. (1998). Analysis of CAG/CTG repeat size in Chinese subjects with schizophrenia and bipolar affective disorder using the repeat expansion detection method. Biological Psychiatry *44*, 1160-1165.

Lindblad, K., Nylander, P. O., De bruyn, A., Sourey, D., Zander, C., Engstrom, C., Holmgren, G., Hudson, T., Chotai, J., Mendlewicz, J., Vanbroeckhoven, C., Schalling, M., and Adolfsson, R. (1995). Detection of expanded CAG repeats in bipolar affective disorder using the repeat expansion detection (RED) method. Neurobiological Disorders *2*, 55-62.

Lindblad, K., Nylander, P. O., Zander, C., Yuan, Q. P., Stahle, L., Engstrom, C., Balciuniene, J., Pettersson, U., Breschel, T., McInnis, M., Ross, C. A., Adolfsson, R., and Schalling, M. (1998). Two commonly expanded CAG/CTG repeat loci: involvement in affective disorders? Molecular Psychiatry *3*, 405-410.

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

Lipkin, S. M., Wang, V., Jacoby, R., Banerjee-Basu, S., Baxevanis, A. D., Lynch, H. T., Elliott, R. M., and Collins, F. S. (2000). MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability. Nature Genetics *24*, 27-35.

Liu, A. Y., Corey, E., Vessella, R. L., Lange, P. H., True, L. D., Huang, G. M., Nelson, P. S., and Hood, L. (1997). Identification of differentially expressed prostate genes: increased expression of transcription factor ETS-2 in prostate cancer. Prostate *30*, 145-153.

Liu, B., Parsons, R., Papadopoulos, N., Nicolaides, N. C., Lynch, H. T., Watson, P., Jass, J. R., Dunlop, M., Wyllie, A., Peltomaki, P., de la Chapelle, A., Hamilton, S. R., Vogelstein, B., and Kinzler, K. W. (1996). Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients. Nature Medicine *2*, 169-174.

Liu, Y., Ray, S. K., Yang, X. Q., Luntz-Leybman, V., and Chiu, I. M. (1998). A splice variant of E2-2 basic helix-loop-helix protein represses the brain-specific fibroblast growth factor 1 promoter through the binding to an imperfect E-box. Journal of Biological Chemistry *273*, 19269-19276.

MacDonald, M. E., Barnes, G., Srinidhi, J., Duyao, M. P., Ambrose, C. M., Myers, R. H., Gray, J., Conneally, P. M., Young, A., Penney, J., Shoulson, I., Hollingsworth, Z., Koroshetz, W., Bird, E., Vonsattel, J.P., Bonilla, E., Moscowitz, C., Penchaszadeh, G., Brzustowicz, L., Alvir, J., Conde, J.B., Cha, J.H., Dure, L., Gomez, F., Ramosarroyo, M., Sanchezramos, J, Snodgrass, S.R., Deyoung, M., Wexler, N.S., Macfarlane, H., Anderson, M.A., Jenkins, B., and Gusella, J.F. (1993). Gametic but not somatic instability of CAG repeat length in Huntington's disease. Journal of Medical Genetics *30*, 982-986.

Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'Hoy, K., Leblond, S., Earlemacdonald, J., Dejong, P.J., Wieringa, B., and Korneluk, R.G. (1992). Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. Science *255*, 1253-1255.

Maier, W., Hallmayer, J., Zill, P., Bondy, B., Lichtermann, D., Ackenheil, M., Minges, J., and Wildenauer, D. (1995). Linkage analysis between pericentrometric markers on chromosome 18 and bipolar disorder: a replication test. Psychiatry Research *59*, 7-15.

Margolis, R. L., Stine, O. C., McInnis, M. G., Ranen, N. G., Rubinsztein, D. C., Leggo, J., Brando, L. V., Kidwai, A. S., Loev, S. J., Breschel, T. S., Callahan, C., Simpson, S. G., DePaulo, J. R., McMahon, F. J., Jain, S., Paykel, E. S., Walsh, C., DeLisi, L. E., Crow, T. J., Torrey, E. F., Ashworth, R. G., Macke, J. P., Nathans, J., and Ross, C. A. (1996). cDNA cloning of a human homologue of the Caenorhabditis elegans cell fate-determining gene mab-21: expression, chromosomal localization and analysis of a highly polymorphic (CAG)n trinucleotide repeat. Human Molecular Genetics *5*, 607-616.

Martorell, L., Monckton, D. G., Gamez, J., Johnson, K. J., Gich, I., de Munain, A. L., and Baiget, M. (1998). Progression of somatic CTG repeat length heterogeneity in the blood cells of myotonic dystrophy patients. Human Molecular Genetics *7*, 307-312.

Massari, A., Gennarelli, M., Menegazzo, E., Pizzuti, A., Silani, V., Mastrogiacomo, I., Pagani, E., Angelini, C., Scarlato, G., Novelli, G., and Dallapiccola, B. (1995). Postzygotic instability of the myotonic dystrophy p[AGC] in repeat supported by larger expansions in muscle and reduced amplifications in sperm. Journal of Neurology *242*, 379-383.

Matias-Guiu, X., and Prat, J. (1998). Molecular pathology of ovarian carcinomas. Virchows Archives *433*, 103-111.

Matson, S. W., and George, J. W. (1987). DNA helicase II of Escherichia coli. Characterization of the single- stranded DNA-dependent NTPase and helicase activities. Journal of Biological Chemistry *262*, 2066-2076.

Maubert, E., Slama, A., Ciofi, P., Viollet, C., Tramu, G., Dupouy, J. P., and Epelbaum, J. (1994). Developmental patterns of somatostatin-receptors and somatostatin- immunoreactivity during early neurogenesis in the rat. Neuroscience *62*, 317-325.

McKee, B.D., and Borromeo, D. (1998) Drosophila melanogaster mutL homolog (Mlh1) gene, complete cDNA sequence. Accession No. AF068257 .

McInnes, L. A., Escamilla, M. A., Service, S. K., Reus, V. I., Leon, P., Silva, S., Rojas, E., Spesny, M., Baharloo, S., Blankenship, K., Peterson, A., Tyler, D., Shimayoshi, N., Tobey, C., Batki, S., Vinogradov, S., Meza, L., Gallegos, A., Fournier, E., Smith, L. B., Barondes, S. H., Sandkuijl, L. A., and Freimer, N. B. (1996). A complete genome screen for genes predisposing to severe bipolar disorder in two Costa Rican pedigrees. Proceedings of the National Academy of Sciences U S A *93*, 13060-13065.

McInnis, M. G., and Margolis, R. L. (1998). Anticipation and Psychiatric Disorders. Genetic Instabilities and Hereditary Neurological Diseases, Academic Press, 401-408.

McMahon, F. J., Hopkins, P. J., Xu, J., McInnis, M. G., Shaw, S., Cardon, L., Simpson, S. G., MacKinnon, D. F., Stine, O. C., Sherrington, R., Meyers, D. A., and DePaulo, J. R. (1997). Linkage of bipolar affective disorder to chromosome 18 markers in a new pedigree series. American Journal of Human Genetics *61*, 1397-1404.

McManus, D. T., Yap, E. P., Maxwell, P., Russell, S. E., Toner, P. G., and McGee, J. O. (1994). p53 expression, mutation, and allelic deletion in ovarian cancer. Journal of Pathology *174*, 159-168.

Mendlewicz, J., Lindbald, K., Souery, D., Mahieu, B., Nylander, P. O., De Bruyn, A., Zander, C., Engstrom, C., Adolfsson, R., Van Broeckhoven, C., Schalling, M., and Lipp, O. (1997). Expanded trinucleotide CAG repeats in families with bipolar affective disorder. Biological Psychiatry *42*, 1115-1122.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L. M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., Mcclure, M., Frye, C., Hattier, T., Phelps, R., Haugenstrano, A., Katcher, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bogden, R., Dayananth, P., Ward, J, Tonin, P., Narod, S., Bristow, P.K., Norris, F.H., Helvering, L., Morrison, P., Rosteck, P., Lai, M., Barrett, J.C., Lewis, C., Neuhausen, S., Cannonalbright, L., Goldgar, D., Wiseman, R., Kamb, A., and Skolnick, M.H. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science *266*, 66-71.

Milner, B. J., Allan, L. A., Eccles, D. M., Kitchener, H. C., Leonard, R. C., Kelly, K. F., Parkin, D. E., and Haites, N. E. (1993). p53 mutation is a common genetic event in ovarian carcinoma. Cancer Research *53*, 2128-2132.

Miyaki, M., Konishi, M., Tanaka, K., Kikuchi-Yanoshita, R., Muraoka, M., Yasuno, M., Igari, T., Koike, M., Chiba, M., and Mori, T. (1997). Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. Nature Genetics *17*, 271-272.

Modrich, P. (1991). Mechanisms and biological effects of mismatch repair. Annual Review of Genetics *25*, 229-253.

Modrich, P. (1989). Methyl-directed DNA mismatch correction. Journal of Biological Chemistry *264*, 6597-6600.

Moises, H. W., Yang, L., Kristbjarnarson, H., Wiese, C., Byerley, W., Macciardi, F., Arolt, V., Blackwood, D., Liu, X., Sjogren, B., Aschauer, H.N., Hwu, H.G., Jang, K., Livesley, W.J., Kennedy, J.L., Zoega, T., Ivarsson, O., Bui, M.T., Yu, M.H., Havsteen, B., Commenges, D., Weissenbach, J., Schwinger, E., Gottesman, J.I., Pakstis, A.J., Wetterberg, L., Kidd, K.K., and Helgason, T. (1995). An international two-stage genome-wide search for schizophrenia susceptibility genes. Nature Genetics *11*, 321-324.

Monckton, D. G., Cayuela, M. L., Gould, F. K., Brock, G. J., Silva, R., and Ashizawa, T. (1999). Very large (CAG)(n) DNA repeat expansions in the sperm of two spinocerebellar ataxia type 7 males. Human Molecular Genetics *8*, 2473-2478.

Monckton, D. G., Wong, L. J., Ashizawa, T., and Caskey, C. T. (1995). Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. Human Molecular Genetics *4*, 1-8.

Mornet, E., Chateau, C., Hirst, M. C., Thepot, F., Taillandier, A., Cibois, O., and Serre, J. L. (1996). Analysis of germline variation at the FMR1 CGG repeat shows variation in the normal-premutated borderline range. Human Molecular Genetics *5*, 821-825.

Morris, A. G., Gaitonde, E., McKenna, P. J., Mollon, J. D., and Hunt, D. M. (1995). CAG repeat expansions and schizophrenia: association with disease in females and with early age-at-onset. Human Molecular Genetics *4*, 1957-1961.

Nakamoto, M., Takebayashi, H., Kawaguchi, Y., Narumiya, S., Taniwaki, M., Nakamura, Y., Ishikawa, Y., Akiguchi, I., Kimura, J., and Kakizuka, A. (1997). A CAG/CTG expansion in the normal population. Nature Genetics *17*, 385-386.

Nancarrow, J. K., Kremer, E., Holman, K., Eyre, H., Doggett, N. A., Le Paslier, D., Callen, D. F., Sutherland, G. R., and Richards, R. I. (1994). Implications of FRA16A structure for the mechanism of chromosomal fragile site genesis. Science *264*, 1938-1941.

Nelson-Piercy, C., Hammond, P. J., Gwilliam, M. E., Khandan-Nia, N., Myers, M. J., Ghatei, M. A., and Bloom, S. R. (1994). Effect of a new oral somatostatin analog (SDZ CO 611) on gastric emptying, mouth to cecum transit time, and pancreatic and gut hormone release in normal male subjects. Journal of Clinical Endocrinological Metabolism *78*, 329-336.

Nicolaides, N. C., Papadopoulos, N., Liu, B., Wei, Y. F., Carter, K. C., Ruben, S. M., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M., Adams, M.D., Venter, J.C., Dunlop, M.G., Hamilton, S.R., Petersen, G.M., Delachapelle, A., Vogelstein, B., Kinzler, and K.W. (1994). Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature *371*, 75-80.

Nielsen, J. E., Koefoed, P., Abell, K., Hasholt, L., Eiberg, H., Fenger, K., Niebuhr, E., and Sorensen, S. A. (1997). CAG repeat expansion in autosomal dominant pure spastic paraplegia linked to chromosome 2p21-p24. Human Molecular Genetics *6*, 1811-1816.

Nikiforov, Y. E., Nikiforova, M., and Fagin, J. A. (1998). Prevalence of minisatellite and microsatellite instability in radiation- induced post-Chernobyl pediatric thyroid carcinomas. Oncogene *17*, 1983-1988.

Nolin, S. L., Houck, G. E., Jr., Gargano, A. D., Blumstein, H., Dobkin, C. S., and Brown, W. T. (1999). FMR1 CGG-repeat instability in single sperm and lymphocytes of fragile- X premutation males. American Journal of Human Genetics *65*, 680-688.

Nystrom-Lahti, M., Wu, Y., Moisio, A. L., Hofstra, R. M., Osinga, J., Mecklin, J. P., Jarvinen, H. J., Leisti, J., Buys, C. H., de la Chapelle, A., and Peltomaki, P. (1996). DNA mismatch repair gene mutations in 55 kindreds with verified or putative hereditary non-polyposis colorectal cancer. Human Molecular Genetics *5*, 763-769.

O'Donovan, M. C., Guy, C., Craddock, N., Murphy, K. C., Cardno, A. G., Jones, L. A., Owen, M. J., and McGuffin, P. (1995). Expanded CAG repeats in schizophrenia and bipolar disorder. Nature Genetics *10*, 380-381.

O'Donovan, M. C., and Owen, M. J. (1999). Candidate-gene association studies of schizophrenia. American Journal of Human Genetics *65*, 587-592.

Offit, K. (1996). Breast cancer and BRCA1 mutations. New England Journal of Medicine *334*, 1197-8; discussion 1199-1200.

Orr, H. T., Chung, M. Y., Banfi, S., Kwiatkowski, T. J., Jr., Servadio, A., Beaudet, A. L., McCall, A. E., Duvick, L. A., Ranum, L. P., and Zoghbi, H. Y. (1993). Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nature Genetics *4*, 221-226.

Oruc, L., Lindblad, K., Verheyen, G. R., Ahlberg, S., Jakovljevic, M., Ivezic, S., Raeymaekers, P., Van Broeckhoven, C., and Schalling, M. (1997). CAG repeat expansions in bipolar and unipolar disorders. American Journal of Human Genetics *60*, 730-732.

Papadopoulos, N., Nicolaides, N. C., Wei, Y. F., Ruben, S. M., Carter, K. C., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M., Adams, M. D., Venter, J.C., Hamilton, S.R., Petersen, G.M., Watson, P., Lynch, H.T., Peltomaki, P., Mecklin, J.P., Delachapelle, A., Kinzler, K.W., and Vogelstein, B. (1994). Mutation of a mutL homolog in hereditary colon cancer. Science *263*, 1625-1629.

Park, T. W., Felix, J. C., and Wright, T. C., Jr. (1995). X chromosome inactivation and microsatellite instability in early and advanced bilateral ovarian carcinomas. Cancer Research *55*, 4793-4796.

Parrish, J. E., Oostra, B. A., Verkerk, A. J., Richards, C. S., Reynolds, J., Spikes, A. S., Shaffer, L. G., and Nelson, D. L. (1994). Isolation of a GCC repeat showing expansion in FRAXF, a fragile site distal to FRAXA and FRAXE. Nature Genetics *8*, 229-235.
Parsons, R., Li, G. M., Longley, M., Modrich, P., Liu, B., Berk, T., Hamilton, S. R., Kinzler, K. W., and Vogelstein, B. (1995). Mismatch repair deficiency in phenotypically normal human cells. Science *268*, 738-740.

Pauls, D. L., Ott, J., Paul, S. M., Allen, C. R., Fann, C. S., Carulli, J. P., Falls, K. M., Bouthillier, C. A., Gravius, T. C., Keith, T. P., Egeland, J.A., and Ginns, E.I. (1995). Linkage analyses of chromosome 18 markers do not identify a major susceptibility locus for bipolar affective disorder in the Old Order Amish. American Journal of Human Genetics *57*, 636-643.

Pearson, C. E., and Sinden, R. R. (1998). Slipped Strand DNA, Dynamic Mutations, and Human Disease. Genetic Instabilities and Hereditary Neurological Diseases, Academic Press, 585-616.

Pedroni, M., Tamassia, M. G., Percesepe, A., Roncucci, L., Benatti, P., Lanza, G., Jr., Gafa, R., Di Gregorio, C., Fante, R., Losi, L., Gallinari, L., Scorcioni, F., Vaccina, F., Rossi, G., Cesinaro, A. M., and Ponz de Leon, M. (1999). Microsatellite instability in multiple colorectal tumors. International Journal of Cancer *81*, 1-5.

Peltomaki, P., Aaltonen, L. A., Sistonen, P., Pylkkanen, L., Mecklin, J. P., Jarvinen, H., Green, J. S., Jass, J. R., Weber, J. L., Leach, F. S., Petersen, G.M., Hamilton, S.R., Delachapelle, A., and Vogelstein, B. (1993). Genetic mapping of a locus predisposing to human colorectal cancer. Science *260*, 810-812.

Phillips, N., Ziegler, M., Saha, B., and Xynos, F. (1993). Allelic loss on chromosome 17 in human ovarian cancer. International Journal of Cancer *54*, 85-91.

Phillips, N. J., Ziegler, M. R., Radford, D. M., Fair, K. L., Steinbrueck, T., Xynos, F. P., and Donis-Keller, H. (1996). Allelic deletion on chromosome 17p13.3 in early ovarian cancer. Cancer Research *56*, 606-611.

Pieretti, M., Cavalieri, C., Conway, P. S., Gallion, H. H., Powell, D. E., and Turker, M. S. (1995). Genetic alterations distinguish different types of ovarian tumors. International Journal of Cancer *64*, 434-440.

Prolla, T. A., Christie, D. M., and Liskay, R. M. (1994). Dual requirement in yeast DNA mismatch repair for MLH1 and PMS1, two homologs of the bacterial mutL gene. Molecular and Cellular Biology *14*, 407-415.

Pscherer, A., Dorflinger, U., Kirfel, J., Gawlas, K., Ruschoff, J., Buettner, R., and Schule, R. (1996). The helix-loop-helix transcription factor SEF-2 regulates the activity of a novel initiator element in the promoter of the human somatostatin receptor II gene. Embo Journal *15*, 6680-6690.

Pulst, S. M., Nechiporuk, A., Nechiporuk, T., Gispert, S., Chen, X. N., Lopes-Cendes, I., Pearlman, S., Starkman, S., Orozco-Diaz, G., Lunkes, A., DeJong, P., Rouleau, G. A., Auburger, G., Korenberg, J. R., Figueroa, C., and Sahba, S. (1996). Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. Nature Genetics *14*, 269-276.

Pulver, A. E., Karayiorgou, M., Lasseter, V. K., Wolyniec, P., Kasch, L., Antonarakis, S., Housman, D., Kazazian, H. H., Meyers, D., and Nestadt, G. (1994). Follow-up of a report of a potential linkage for schizophrenia on chromosome 22q12-q13.1: Part 2. American Journal of Medical Genetics *54*, 44-50.

Pulver, A. E., Lasseter, V. K., Kasch, L., Wolyniec, P., Nestadt, G., Blouin, J. L., Kimberland, M., Babb, R., Vourlis, S., Chen, H., Lalioti, M., Morris, M.A., Karayiorgou, M., Ott, J., Meyers, D., Antonarakis, S.E., Housman, D., and Kazazian, H.H. (1995). Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. American Journal of Medical Genetics *60*, 252-260.

Rens-Domiano, S., and Reisine, T. (1992). Biochemical and functional properties of somatostatin receptors. Journal of Neurochemistry *58*, 1987-1996.

Riess, O., Epplen, J. T., Amoiridis, G., Przuntek, H., and Schols, L. (1997). Transmission distortion of the mutant alleles in spinocerebellar ataxia. Human Genetics *99*, 282-284.

Rodabaugh, K. J., Biggs, R. B., Qureshi, J. A., Barrett, A. J., Welch, W. R., Bell, D. A., Berkowitz, R. S., and Mok, S. C. (1995). Detailed deletion mapping of chromosome 9p and p16 gene alterations in human borderline and invasive epithelial ovarian tumors. Oncogene *11*, 1249-1254.

Rodabaugh, K. J., Blanchard, G., Welch, W. R., Bell, D. A., Berkowitz, R. S., and Mok, S. C. (1995). Detailed deletion mapping of chromosome 6q in borderline epithelial ovarian tumors. Cancer Research *55*, 2169-2172.

Rubinsztein, D. C., Leggo, J., Crow, T. J., DeLisi, L. E., Walsh, C., Jain, S., and Paykel, E. S. (1996). Analysis of polyglutamine-coding repeats in the TATA-binding protein in different human populations and in patients with schizophrenia and bipolar affective disorder. American Journal of Medical Genetics *67*, 495-8.

Sanpei, K., Takano, H., Igarashi, S., Sato, T., Oyake, M., Sasaki, H., Wakisaka, A., Tashiro, K., Ishida, Y., Ikeuchi, T., Koide, R., Saito, M., Sato, A., Tanaka, T., Hanyu, S., Takiyama, Y., Nishizawa, M., Shimizu, N., Nomura, Y., Segawa, M., Iwabuchi, K., Eguchi, I., Tanaka, H., Takahashi, H., and Tsuji, S. (1996). Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. Nature Genetics *14*, 277-284.

Saretzki, G., Hoffmann, U., Rohlke, P., Psille, R., Gaigal, T., Keller, G., Hofler, H., Loning, T., Petersen, I., and Dietel, M. (1997). Identification of allelic losses in benign, borderline, and invasive epithelial ovarian tumors and correlation with clinical outcome. Cancer *80*, 1241-1249.

Schalling, M., Hudson, T. J., Buetow, K. H., and Housman, D. E. (1993). Direct detection of novel expanded trinucleotide repeats in the human genome. Nature Genetics *4*, 135-139.

Schraen-Maschke, S., Brique, S., Chartier-Harlin, M. C., Brique, E., Destee, A., and Sablonniere, B. (1999). Analysis of ERDA1, CTG18.1, and uncloned CAG/CTG repeat sequences in familial Parkinson's disease with anticipation [In Process Citation]. American Journal of Medical Genetics *88*, 738-741.

Shih, Y. C., Kerr, J., Hurst, T. G., Khoo, S. K., Ward, B. G., and Chenevix-Trench, G. (1998). No evidence for microsatellite instability from allelotype analysis of benign and low malignant potential ovarian neoplasms. Gynecological Oncology *69*, 210-213.

Sidransky, E., Burgess, C., Ikeuchi, T., Lindblad, K., Long, R. T., Philibert, R. A., Rapoport, J., Schalling, M., Tsuji, S., and Ginns, E. I. (1998). A triplet repeat on 17q accounts for most expansions detected by the repeat-expansion-detection technique. American Journal of Human Genetics *62*, 1548-1551.

Skerjanc, I. S., Truong, J., Filion, P., and McBurney, M. W. (1996). A splice variant of the ITF-2 transcript encodes a transcription factor that inhibits MyoD activity. Journal of Biological Chemistry *271*, 3555-3561.

Slamon, D. J., Godolphin, W., Jones, L. A., Holt, J. A., Wong, S. G., Keith, D. E., Levin, W. J., Stuart, S. G., Udove, J., Ullrich, A., and Press, M.F. (1989). Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science *244*, 707-712.

Stine, O. C., Xu, J., Koskela, R., McMahon, F. J., Gschwend, M., Friddle, C., Clark, C. D., McInnis, M. G., Simpson, S. G., Breschel, T. S., Vishio, E., Riskin, K., Feilotter, H., Chen, E., Shen, S., Folstein, S., Meyers, D.A., Botstein, D., Marr, T.G., and Depaulo, J.R. (1995). Evidence for linkage of bipolar disorder to chromosome 18 with a parent- of-origin effect. American Journal of Human Genetics *57*, 1384-1394.

Stober, G., Jatzke, S., Meyer, J., Okladnova, O., Knapp, M., Beckmann, H., and Lesch, K. P. (1998). Short CAG repeats within the hSKCa3 gene associated with schizophrenia: results of a family-based study. Neuroreport *9*, 3595-3599.

Strathdee, G., MacKean, M. J., Illand, M., and Brown, R. (1999). A role for methylation of the hMLH1 promoter in loss of hMLH1 expression and drug resistance in ovarian cancer. Oncogene *18*, 2335-2341.

Straub, R. E., Lehner, T., Luo, Y., Loth, J. E., Shao, W., Sharpe, L., Alexander, J. R., Das, K., Simon, R., Fieve, R. R., Lerer, B., Endicott, J., Ott, J., Gilliam, T.C., and Baron, M. (1994). A possible vulnerability locus for bipolar affective disorder on chromosome 21q22.3. Nature Genetics *8*, 291-296.

Straub, R. E., MacLean, C. J., O'Neill, F. A., Walsh, D., and Kendler, K. S. (1997). Support for a possible schizophrenia vulnerability locus in region 5q22- 31 in Irish families. Molecular Psychiatry *2*, 148-155.

Strauss, J., Zhang, X. R., Barron, Y., Ganguli, R., and Nimgaonkar, V. L. (1999). Lack of association between schizophrenia and a pancreatic phospholipase A-2 gene (PLA2G1B) polymorphism. Psychiatric Genetics *9*, 153-155.

Su, S. S., and Modrich, P. (1986). Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs. Proceedings of the National Academy of Sciences U S A *83*, 5057-5061.

Takahashi, H., Chiu, H. C., Bandera, C. A., Behbakht, K., Liu, P. C., Couch, F. J., Weber, B. L., LiVolsi, V. A., Furusato, M., Rebane, B. A., Cardonick, A., Benjamin, I., Morgan, M. A., King, S. A., Mikuta, J. J., Rubin, S. C., and Boyd, J. (1996). Mutations of the BRCA2 gene in ovarian carcinomas. Cancer Research *56*, 2738-2741.

Takiyama, Y., Sakoe, K., Amaike, M., Soutome, M., Ogawa, T., Nakano, I., and Nishizawa, M. (1999). Single sperm analysis of the CAG repeats in the gene for dentatorubral-pallidoluysian atrophy (DRPLA): the instability of the CAG repeats in the DRPLA gene is prominent among the CAG repeat diseases. Human Molecular Genetics *8*, 453-457.

Takiyama, Y., Sakoe, K., Soutome, M., Namekawa, M., Ogawa, T., Nakano, I., Igarashi, S., Oyake, M., Tanaka, H., Tsuji, S., and Nishizawa, M. (1997). Single sperm analysis of the CAG repeats in the gene for Machado-Joseph disease (MJD1): evidence for non-Mendelian transmission of the MJD1 gene and for the effect of the intragenic CGG/GGG polymorphism on the intergenerational instability. Human Molecular Genetics *6*, 1063-1068.

Tanaka, F., Reeves, M. F., Ito, Y., Matsumoto, M., Li, M., Miwa, S., Inukai, A., Yamamoto, M., Doyu, M., Yoshida, M., Hashizume, Y., Terao, S., Mitsuma, T., and Sobue, G. (1999). Tissue-Specific Somatic Mosaicism in Spinal and Bulbar Muscular Atrophy Is Dependent on CAG-Repeat Length and Androgen Receptor-Gene Expression Level. American Journal of Human Genetics *65*, 966-973.

Tangir, J., Loughridge, N. S., Berkowitz, R. S., Muto, M. G., Bell, D. A., Welch, W. R., and Mok, S. C. (1996). Frequent microsatellite instability in epithelial borderline ovarian tumors. Cancer Research *56*, 2501-2505.

Tavtigian, S. V., Simard, J., Rommens, J., Couch, F., Shattuck-Eidens, D., Neuhausen, S., Merajver, S., Thorlacius, S., Offit, K., Stoppa-Lyonnet, D., Belanger, C., Bell, R., Berry, S., Bogden, R., Chen, Q., Davis, T., Dumont, M., Frye, C., Hattier, T., Jammulapati, S., Janecki, T., Jiang, P., Kehrer, R., Leblanc, J. F., Mitchell, J.T., Mcarthurmorrison, J., Nguyen, K., Peng, Y., Samson, C., Schroeder, M., Snyder, S.C., Steele, L., Stringfellow, M., Stroup, C., Swedlund, B., Swensen, J., Teng, D., Thomas, A., Tran, T., Tranchant, M., Weaverfeldhaus, J., Wong, A.K.C., Shizuya, H., Eyfjord, J.E., Cannonalbright, L., Labrie, F., Skolnick, M.H., Weber, B., Kamb, A., and Goldgar, D. E. (1996). The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. Nature Genetics *12*, 333-337.

Teague, J. W., Morton, N. E., Dennis, N. R., Curtis, G., McKechnie, N., Macpherson, J. N., Murray, A., Pound, M. C., Sharrock, A. J., Youings, S. A., and Jacobs, P. A. (1998). FRAXA and FRAXE: evidence against segregation distortion and for an effect of intermediate alleles on learning disability. Proceedings of the National Academy of Sciences U S A *95*, 719-724.

Telenius, H., Kremer, B., Goldberg, Y. P., Theilmann, J., Andrew, S. E., Zeisler, J., Adam, S., Greenberg, C., Ives, E. J., Clarke, L. A., and Hayden, M.R. (1994). Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. Nature Genetics *6*, 409-414.

Thibaut, F., Martinez, M., Petit, M., Jay, M., and Campion, D. (1995). Further evidence for anticipation in schizophrenia. Psychiatry Research *59*, 25-33.

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

Thibodeau, S. N., French, A. J., Cunningham, J. M., Tester, D., Burgart, L. J., Roche, P. C., McDonnell, S. K., Schaid, D. J., Vockley, C. W., Michels, V. V., Farr, G. H., Jr., and O'Connell, M. J. (1998). Microsatellite instability in colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1. Cancer Research *58*, 1713-1718.

Thornton, C. A., Johnson, K., and Moxley, R. T., 3rd (1994). Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. Annals of Neurology *35*, 104-107.

Umar, A., Boyer, J. C., Thomas, D. C., Nguyen, D. C., Risinger, J. I., Boyd, J., Ionov, Y., Perucho, M., and Kunkel, T. A. (1994). Defective mismatch repair in extracts of colorectal and endometrial cancer cell lines exhibiting microsatellite instability. Journal of Biological Chemistry *269*, 14367-14370.

Vasen, H. F., Mecklin, J. P., Khan, P. M., and Lynch, H. T. (1991). The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum *34*, 424-425.

Vasen, H. F., Watson, P., Mecklin, J. P., and Lynch, H. T. (1999). New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology *116*, 1453-1456.

Verheyen, G. R., Del-Favero, J., Mendlewicz, J., Lindblad, K., Van Zand, K., Aalbregtse, M., Schalling, M., Souery, D., and Van Broeckhoven, C. (1999). Molecular interpretation of expanded RED products in bipolar disorder by CAG/CTG repeats located at chromosomes 17q and 18q. Neurobiological Disorders *6*, 424-432.

Vincent, J. B., Kalsi, G., Klempan, T., Tatuch, Y., Sherrington, R. P., Breschel, T., McInnis, M. G., Brynjolfsson, J., Petursson, H., Gurling, H. M., Gottesman, II, Torrey, E. F., Petronis, A., and Kennedy, J. L. (1998). No evidence of expansion of CAG or GAA repeats in schizophrenia families and monozygotic twins. Human Genetics *103*, 41-47.

Vincent, J. B., Klempan, T., Parikh, S. S., Sasaki, T., Meltzer, H. Y., Sirugo, G., Cola, P., Petronis, A., and Kennedy, J. L. (1996). Frequency analysis of large CAG/CTG trinucleotide repeats in schizophrenia and bipolar affective disorder. Molecular Psychiatry *1*, 141-148.

Vincent, J. B., Kovacs, M., Krol, R., Barr, C. L., and Kennedy, J. L. (1999). Intergenerational CAG repeat expansion at ERDA1 in a family with childhood-onset depression, schizoaffective disorder, and recurrent major depression. American Journal of Medical Genetics *88*, 79-82.

Vincent, J. B., Petronis, A., Strong, E., Parikh, S. V., Meltzer, H. Y., Lieberman, J., and Kennedy, J. L. (1999). Analysis of genome-wide CAG/CTG repeats, and at SEF2-1B and ERDA1 in schizophrenia and bipolar affective disorder. Molecular Psychiatry *4*, 229-234.

Wang, S., Sun, C. E., Walczak, C. A., Ziegle, J. S., Kipps, B. R., Goldin, L. R., and Diehl, S. R. (1995). Evidence for a susceptibility locus for schizophrenia on chromosome 6pter-p22. Nature Genetics *10*, 41-46.

Welsh, K. M., Lu, A. L., Clark, S., and Modrich, P. (1987). Isolation and characterization of the Escherichia coli mutH gene product. Journal of Biological Chemistry *262*, 15624-15629.

Wertheim, I., Muto, M. G., Welch, W. R., Bell, D. A., Berkowitz, R. S., and Mok, S. C. (1994). p53 gene mutation in human borderline epithelial ovarian tumors. Journal of the National Cancer Institute *86*, 1549-1551.

Wertheim, I., Tangir, J., Muto, M. G., Welch, W. R., Berkowitz, R. S., Chen, W. Y., and Mok, S. C. (1996). Loss of heterozygosity of chromosome 17 in human borderline and invasive epithelial ovarian tumors. Oncogene *12*, 2147-2153.

Wheeler, J. M., Beck, N. E., Kim, H. C., Tomlinson, I. P., Mortensen, N. J., and Bodmer, W. F. (1999). Mechanisms of inactivation of mismatch repair genes in human colorectal cancer cell lines: the predominant role of hMLH1. Proceedings of the National Academy of Sciences U S A *96*, 10296-10301.

Wijnen, J., Khan, P. M., Vasen, H., van der Klift, H., Mulder, A., van Leeuwen-Cornelisse, I., Bakker, B., Losekoot, M., Moller, P., and Fodde, R. (1997). Hereditary nonpolyposis colorectal cancer families not complying with the Amsterdam criteria show extremely low frequency of mismatch-repair- gene mutations. American Journal of Human Genetics *61*, 329-335.

Wijnen, J., van der Klift, H., Vasen, H., Khan, P. M., Menko, F., Tops, C., Meijers Heijboer, H., Lindhout, D., Moller, P., and Fodde, R. (1998). MSH2 genomic deletions are a frequent cause of HNPCC. Nature Genetics *20*, 326-328.

Williams GW, Woollard PM, Hingamp P. (1998) "NIX: A Nucleotide Identification system at the HGMP-RC" URL: http://www.hgmp.mrc.ac.uk/NIX/

Wiper, D. W., Zanotti, K. M., Kennedy, A. W., Belinson, J. L., and Casey, G. (1998). Analysis of allelic imbalance on chromosome 17p13 in stage I and stage II epithelial ovarian cancers. Gynecological Oncology *71*, 77-82.

Wittekindt, O., Schwab, S. G., Burgert, E., Knapp, M., Albus, M., Lerer, B., Hallmayer, J., Rietschel, M., Segman, R., Borrmann, M., Lichtermann, D., Crocq, M. A., Maier, W., Morris-Rosendahl, D. J., and Wildenauer, D. B. (1999). Association between hSKCa3 and schizophrenia not confirmed by transmission disequilibrium test in 193 offspring/parents trios. Molecular Psychiatry *4*, 267-270.

Wohrle, D., Kennerknecht, I., Wolf, M., Enders, H., Schwemmle, S., and Steinbach, P. (1995). Heterogeneity of DM kinase repeat expansion in different fetal tissues and further expansion during cell proliferation in vitro: evidence for a casual involvement of methyl-directed DNA mismatch repair in triplet repeat stability. Human Molecular Genetics *4*, 1147-1153.

Wong, L. J., Ashizawa, T., Monckton, D. G., Caskey, C. T., and Richards, C. S. (1995). Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. American Journal of Human Genetics *56*, 114-122.

Worsley, S. D., Ponder, B. A., and Davies, B. R. (1997). Overexpression of cyclin D1 in epithelial ovarian cancers. Gynecological Oncology *64*, 189-195.

Wu, Y., Berends, M. J., Mensink, R. G., Kempinga, C., Sijmons, R. H., van Der Zee, A. G., Hollema, H., Kleibeuker, J. H., Buys, C. H., and Hofstra, R. M. (1999). Association of hereditary nonpolyposis colorectal cancer-related tumors displaying low microsatellite instability with MSH6 germline mutations. American Journal of Human Genetics *65*, 1291-1298.

Zhang, L., Leeflang, E. P., Yu, J., and Arnheim, N. (1994). Studying human mutations by sperm typing: instability of CAG trinucleotide repeats in the human androgen receptor gene. Nature Genetics *7*, 531-535.

Zhuang, Y., Cheng, P., and Weintraub, H. (1996). B-lymphocyte development is regulated by the combined dosage of three basic helix-loop-helix genes, E2A, E2-2, and HEB. Molecular and Cellular Biology *16*, 2898-2905.

Zhuchenko, O., Bailey, J., Bonnen, P., Ashizawa, T., Stockton, D. W., Amos, C., Dobyns, W. B., Subramony, S. H., Zoghbi, H. Y., and Lee, C. C. (1997). Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. Nature Genetics *15*, 62-69.

Zuhlke, C., Riess, O., Bockel, B., Lange, H., and Thies, U. (1993). Mitotic stability and meiotic variability of the (CAG)n repeat in the Huntington disease gene. Human Molecular Genetics *2*, 2063-2067.

Appendix One

CEPH DNA Samples

Sample Name	*CTG18.1*
F-01 1332	11,11
M-02 1332	11,52
F-01 1334	17,25
M-02 1334	17,29
F-01 1344	11,11
M-02 1344	17,118
F-01 1420	15,330
M-02 1420	11,17
F-01 1424	11,26
M-02 1424	14,16

Radiotherapy Sperm DNA Samples

Sample Name	*CTG18.1*	*ERDA1*	*DM-1*
1580	24,29	21,21	5,32
1699	11,11	10,21	5,5
1796	11,14	10,22	5,12
1840	11,17	14,65	5,13
1868	11,11	17,22	5,13
1969	11,29	10,52	12,29
1970	11,31	22,22	5,34
2038	14,14	10,22	13,14
2044	17,17	10,59	5,5
2048	11,17	16,23	5,5

DM-1 Blood DNA

Sample Name	*CTG18.1*	*ERDA1*
AJ	17,26	20,53
AS	11,14	11,20
AVM	14,25	20,56
CM	17,24	21,69
DF	14,32	10,20
DM	17,18	17,21
EM	17	10,17
JM	14,23	17,21
KW	11,28	17,19
MB	14,25	17,21
MJ	17	10,68
MS	11,26	10,19
MW	11,17	21,57
NJ	11,15	21,22
PaW	21,26	10,20
PeW	11,15	10,19
RG	14,22	10,20
RJ	11,15	22,61
RM	18,79	21,68
RS	11,14	20,21
SN	14,16	20
TB	17,22	10
TP	17,69	10
VW	17,54	10,21
WW	24,25	20
YM	11,72	10,21

DM-1 patient blood DNA

Sample Name	*CTG18.1*	*ERDA1*	Sample Name	*CTG18.1*	*ERDA1*
101-3001	11,17	10,21	125-3011	11,27	21,33
101-3004	11,17	21,69	125-4014	11,17	21,23
101-3008	11	10,21	137-1538	34,35	10,55
101-3056	15,69	10,22	137-2004	15,35	22,54
101-3057	17,68	22	137-2664	15,35	10,22
101-3059	11,15	10,22	137-3666	15	10,21/22
101-3061	11,17	10,22	152-2007	11,17	10,21
101-3065	15,17	22,59	152-3021	27,89	21,21
101-3071	11,30	22,72	152-3633	11,17	10,17
101-3080	11,30	21,68	164-2009	11	10,21
101-3082	11,30	21,71	164-3015	11	10,21
101-4001	11	10,22	175-1001	11,25	10,22
101-4004	11,51	10,21	175-2007	25,26	10,22
101-4010	11	10,22	175-2009	17,25	10,17
101-4033	11,28	10,10	175-2012	11,26	10,22
101-4035	11,28	10,10			
101-4174	12,15	10,59			
101-4177	11,69	10,22			
101-4180	15,17	10,22			
101-4184	14,15	10,22			
101-4205	15,29	17,73			
101-4221	11,17	21,62			
101-4222	17,30	21,64			
101-4223	11,17	21,61			
101-5025	18	21,22			
101-5053	11,28	10,62			
114-1001	11,25	10,21			
114-2005	25,34	20,21			
114-3008	11,34	20,63			
114-3673	11,25	21,21			
119-2004	11,26	20,22			
119-2005	17	20,22			
119-3011	17,26	10,22			
119-3017	17,25	17,22			
125-2016	17,25	10,15			

Ovarian Tumours and Cell line DNA

Sample Name	*CTG18.1*	*ERDA1*	*DM-1*
LOVO	10, 13	10, 13	9, 17
AN3CA	5	11	10, 20
A2780	11,13	16, 27	10
MCP1	11, 13	16, 27	10
MCP2	11, 13	16, 28	10
MCP3	11, 13	16, 28	10
MCP5	11, 13	16, 26	10
MCP6	11, 13	16, 28	10
MCP7	11, 13	16, 27	10
MCP8	11, 13	16, 27	10
MCP9	11, 13	16, 28	10
OT437	11, 13	16	10
OT450	5	11	10, 21
OT481	5, 13	17, 106	21, 60
OT503	11, 13	16	10
POU478	10, 13	16	10
POWFLL	11, 29	15,17	21, 27
DH	5, 13	11	21
WR	5, 19	11, 14	20
SC	5, 13	14	10, 21
CC	11, 13	17, 24	10, 21
PM	5, 13	12, 18	10
IM	14, 15	16	16
JC	10, 11	11, 14	10, 22
LD	5, 21	11, 30	10, 60
AC	5, 23	11	10, 21
PK	5, 11	11, 32	10, 53
LC	5, 26	11, 64	22, 51
MT	5, 12	14	10, 68
EH	5, 13	11, 24	10, 21
BD	11, 12	11, 26	52, 62
AM	13	21, 23	12, 19

Chimpanzee DNA

Sample Name	*CTG18.1*	*ERDA1*	*DM-1*
FRITS	11	19, 22	22, 24
LOUISE	11, 12	19, 28	12, 19
ZEEF	11, 12	19, 28	19, 22
MAANTEN	11, 12	19	19, 22
VANESSA	11, 12	22, 28	19, 24
DIANA	11	18, 19	15, 19
FLINT	11	19, 22	15, 24
DORIEN	11	19, 22	15, 22
SHERRY	11, 14	19, 26	20, 25
CLAUDETTE	11	19	22, 25

HNPCC DNA

LabID	Family	Initials	*DM-1*	*CTG18.1*	*ERDA1*
941713	hn5	MD	5,12	11,17	16,21
950068	hn1	VR	20,23	11,25	22,53
950863	hn9	PP	12,13	11,17	16,19
950864	hn9	BF	5,12	17	10,60
951675	hn9	BaF	5,12	16,17	10,60
951816	hn18	EE	5,11	16, 93	21,22
951855	hn18	FT	5,11	11, 93	10,22
952156	hn26	JT	5	11,24	17,66
952505	hn32	MP	15,22	11	10,22
960349	hn44	JC	12,13	17	21,62
964850	hn1	DR	5,23	25,38	20,21
965822	hn9	CP	12	11,22	16,17
968254	hn78	TJ	5	17	10,86
971614	hn86	NS	5,11	11,31	10,75
973409	hn100	GT	13	11	10,17
973411	hn83	CS	13,14	11	10,17
974845	hn83	PC	5,14	11,14	21,22
975563	hn5	PD	12,24	11,17	21,22
976171	hn110	RW	5,8	17,54	10,17
976478	hn5	ID	12,24	18	21,57
976563	hn112	ND	8,16	18,25	17
976652	hn114	PB	5,11	14,17	10,77
977229	hn117	BB	12	15,25	23,58
977510	hn119	ED	12,23	11	21,22
977576	hn100	KT	5,13	11,14	21,55
979006	hn131	CJ	5,12	14,17	21,68
979325	hn134	SC	11	15,17	20,68
986739	hn136	DC	5,12	11,24	10
986742	hn112	WC	13,22	11,15	17
986743	hn112	SS	8,13	14,18	10,17
986750	hn112	ND	8,11	14,25	17,22
988279	hn134	AC	11,13	15	10,20
992777	hn112	AA	13,15	11,15	17
9604517	hn1	TS			10,22
9702797	hn110	BW		16,55	
	hn18	GF			

LabID	Family	Initials	*CTG18.1*	*ERDA1*
000/0211	211	BB	17	19,21
13/3173	3173	BeMa	14,81	10
046/3069	3069	BeMc	11,46	10,16
22/3174	3174	BrM	11	10,15
031/3062	3062	BV	17,24	7,20
14/3038	3038	CB	17,26	20,21
021/3007	3007	CF	11,14	21,59
/3096	3096	CG	11	10,21
008/3166	3166	CK	14,15	10,19
036/3202	3202	CL	11,32	10
020/3173	3173	CR	14,17	10
015/3069	3069	CW	11	16,21
080/3062	3062	DA	11,14	20
000/0276	276	DR	11,24	10,20
012/3065	3065	DT	11,15	15,67
29/3062	3062	EA	11,17	10,22
000/3202	3202	FP	11,25	10,16
000/0155	155	GA	11,17	10,16
000/540	540	JaM	25,27	10,14
/3148	3148	JC	15,27	10
014/3028	3028	JH	17	10,16
026/469	469	JiM	11,16	10,58
136/3001	3001	JiW	13,15	10,19
21/3157	3157	JoW	11,17	10
080/3011	3011	JT	17,23	10,15
000/534	534	KH	11,17	20,21
17/3014	3014	KP	13,17	16,21
000/0321	321	LC	11	10,56
64/3001	3001	LG	13,24	10,19
022/3202	3202	LM	17,25	10,17
014/3011	3011	LT	17,23	10,16
077/3062	3062	MA	22,23	20
022/3033	3033	MM	11,17	28,59
13/3116	3116	MS	14,17	20,22
000/3107	3107	MT	17,27	20
41/3032	3032	NW	14,29	16
001/3011	3011	OL	11,23	15,16
004/3084	3084	PP	11,17	21,51

HNPCC DNA

LabID	Family	Initials	*CTG18.1*	*ERDA1*
000/3151	3151	RR	11	10
000/3190	3190	RC	11,17	10,38
010/3185	3185	RR	11,17	10
007/3095	3095	SE	14,17	10,20
033/3102	3102	SS	11,28	10,21
024/3032	3032	TW	17	14,61
145/3001	3001	VA	14,17	14,19
004/3127	3127	WC	14,35	10,16
011/3032	3032	WW	11,22	20,61

Semen DNA

Sample Name	CTG18.1	ERDA1	Sample Name	CTG18.1	ERDA1
GS1	11,64	11,22	GS39	-	
GS2			GS40	-	
GS3	22,24	10,22	GS41	11,14	22,61
GS4	14,23	10,22	GS42	11,14	15,22
GS5	15,81	10,22	GS43	-	
GS6	14,17	16,62	GS44	11	16,17
GS7	17,23	16,57	GS45	14,26	10,58
GS8	-		GS46	-	
GS9	14,23	10,21	GS47	11,13	21,23
GS10	11,26	10,27	GS48	17	10,21
GS11	11,17	10,15	GS49	-	
GS12	16,87	21,69	GS50	13,14	21,64
GS13	11,24	21	GS51	11,17	10,21
GS14	11,25	10	GS52	-	
GS15	-		GS53	11,17	10,17
GS16	11	17,23	GS54	-	
GS17	11,22	18,21	GS55	11,27	10,22
GS18	17,25	60,64	GS56	-	
GS19	22,32	24,?	GS57	16,17	21,22
GS20	13,62	10,59	GS58	11,14	10,56
GS21	-		GS59	11,14	10
GS22	15	21,22	GS60	-	
GS23	11,14	10	GS61	-	
GS24	11,14	7,61	GS62	-	
GS25	-		GS63	-	
GS26	17,28	10,54	GS64	-	
GS27	11	17,22	GS65	16,17	23,61
GS28	11,30	10,58	GS66	11,37	10,17
GS29	16,31	22,23	GS67	17,32	21,68
GS30	14,24	10,60	GS68		
GS31	-		GS69	11,17	21,22
GS32	-		GS70	24,33	17
GS33	-		GS71	14,26	21,22
GS34	11,24	10	GS72	-	
GS35	15,25	22,58	GS73	-	
GS36	11,14	10,17	GS74	14,17	21,44
GS37	-		GS75	11,23	17,18
GS38	-		GS76	11,17	10,21

Semen DNA

Sample Name	CTG18.1	ERDA1	Sample Name	CTG18.1	ERDA1
GS77	14,60	10,21	GS115	11,26	16,60
GS78	11,31	15,63	GS116	11,21	10,21
GS79			GS117	11,25	10,52
GS80	11	10,59	GS118	17,24	10,58
GS81	25,26	57	GS119	11,25	10,17
GS82	14,21	17,60	GS120	14	18,21
GS83	11,24	10	GS121	17	18,21
GS84			GS122	11,17	18,21
GS85	11,30	10,58	GS123	14,16	18?,21
GS86	11	21,22	GS124	17,67	-
GS87	17,25	10,55	GS125	11,15	18,21
GS88	11	10,57	GS126	14,24	-
GS89	11,14	10,21	GS127	11,17	10?,21?
GS90	11,23	10,16	GS128	11,24	10,21
GS91	11,14	10,15	GS129	11,17	15,17
GS92	11,16	10,22	GS130	31,41	16,21
GS93	11,33	17	GS131	18,79	16,58
GS94	11,14	10,36	GS132	14	10
GS95	14,25	21,22	GS133	14,26	21
GS96	13,14	10,21	GS134	14	10,22
GS97	11,26	17	GS135	28,37	10
GS98	11	10,72	GS136	14	16,23
GS99	11,17	10,21	GS137	11,14	22
GS100	16,22	17	GS138	14,25	10
GS101	12	21,60	GS139	11,14	21,22
GS102	16,24	21,22	GS140	11,14	20
GS103	11,25	10,56	GS141		
GS104	11,20	10,22	GS142	11,14	10,18
GS105	11,30	18,21	GS143	17,32	10,21
GS106	14,17	17,70	GS144	14	22,60
GS107	11,21	17,58	GS145	13,14	21,63
GS108	14,17	10	GS146	11	10,66
GS109	14,24	10	GS147	-	
GS110	11,17	10,74	GS148	28,36	10
GS111	14,28	10	GS149	11,14	21,22
GS112	17	10,22	GS150	17,29	22
GS113	11,17	10,21	GS151	14,15	13,21
GS114	15,17	10,22	GS152	17,91	10,57

Semen DNA

Sample Name	CTG18.1	ERDA1	Sample Name	CTG18.1	ERDA1
GS153	11,40	10,59	GS191	23,26	10,17
GS154	11	20,21	GS192	14,22	10,21
GS155	11	10,22	GS193	11	21
GS156	11,24	10,22	GS194	17	
GS157	11,14	10,22	GS195	14,27	10,22
GS158	11,13	21,27	GS196	14,24	22,65
GS159	14,17	10	GS197	11,26	16,52
GS160	11,15	10,22	GS198	14,20	10
GS161	11	10,22	GS199	11,30	10,22
GS162	11,16	23,60	GS200	11,14	10,23
GS163	26,30	22	GS201	11,27	10?,16?
GS164	11,14	14,26	GS202	24,25	10,22
GS165	11,22	21,23	GS203	11,25	53
GS166	11,71	10	GS204	11,17	21,27
GS167	17,85	17,22	GS205	11,28	10,23
GS168	13	22,64	GS206	14,17	16,17
GS169	11,17	10,21	GS207	11	23,62
GS170	14,23	11,21	GS208	11,14	19
GS171	11	17,65	GS209	14,21	10
GS172	11	-	GS210	11,14	17,21
GS173	15,25	22,55	GS211	14,15	10,63
GS174	11	17,29	GS212	11,15	10,21
GS175	14,23	10,21	GS213	17,32	21,22
GS176	11,17	18,21	GS214	11,17	10,22
GS177	14,17	16,22	GS215	15,20	10,21
GS178	13,14	10	GS216	11	22
GS179	11,16	10,19	GS217	11,14	10,69
GS180	14,29	21,22	GS218	10,16	22,23
GS181	11,24	21,?	GS219	18,22	15,64
GS182	22,25	21,22	GS220	17,22?	15?
GS183	15,26	10,61	GS221	11,17	16,22
GS184	11	22,57	GS222	11	17,22
GS185	11	21	GS223	11,12	21,22
GS186	19,23	10,21	GS224	11,29	10,21
GS187	11,14	10,21	GS225	16,17	10,20
GS188	20,22	10,26	GS226	11,14	10,15
GS189	11	10,53	GS227	11	22
GS190	11,17	17,58	GS228	16	10

Semen DNA

Sample Name	CTG18.1	ERDA1	Sample Name	CTG18.1	ERDA1
GS229	11,14	10	GS267	11,14	10,22
GS230	17,?	50,58	GS268	17	10,22
GS231	16,23	10,22	GS269	11,22	18,21
GS232	11,14	10,66	GS270	11,17	10,16
GS233	11,14	10,66	GS271	11,24	10,21
GS234	24,28	10	GS272	15,33	10,22
GS235	15,24	21,72	GS273	11,75	17,21
GS236	11,17	23,71	GS274	14,16	10
GS237	11	10,22	GS275	11,17	10,17
GS238	17	10,60	GS276	11	10,21
GS239	11,14	10,18	GS277	11,12	21,62
GS240	11,17	21	GS278	11,14	17,65
GS241	14,17	17,21	GS279	14,17	10
GS242	11,17	10	GS280	15,25	21,22
GS243	11,25	10,22	GS281	13,15	10
GS244	11,14	15,17	GS282	14,17	10,21
GS245	11	16,59	GS283	17	10,22
GS246	11	21,22	GS284	11,17	21,25
GS247	11,16	10	GS285	11,24	22
GS248	14,17	16	GS286	17,26	20,49
GS249	11,15	21,58	GS287	11	10
GS250	17	16,20	GS288	24,32	10,55
GS251	17,25	21	GS289	11,17	10,62
GS252	15,26	9	GS290	11,23	17,22
GS253	11,24	10,22	GS291	16,17	21,59
GS254	11	17,21	GS292	11,15	21,27
GS255	11,17	10,15	GS293	11	10,17
GS256	11	22	GS294	11,14	10,59
GS257	17,36	10,16	GS295	14,24	10,23
GS258	14,25	10	GS296	11,20	21,60
GS259	14	17	GS297	-	10,64
GS260	16,31	66	GS298	11,17	17,21
GS261	11,84	10,22	GS299	11,15	10,21
GS262	11,14	10,22	GS300	11,77	17,67
GS263	11,15	21,27	GS301	11,28	10
GS264	14,17	10,60	GS302	14,62	10,21
GS265	11,24	21	GS303	14,64	17,22
GS266	17,24	17,68	GS304	14,17	22,68

Semen DNA

Sample Name	CTG18.1	ERDA1	Sample Name	CTG18.1	ERDA1
GS305	11,14	10	GS343	11,14	59
GS306	11,15	10,82	GS344	11,15	16,20
GS307	11,62	10,21	GS345	14	10,22
GS308	11	10,21	GS346	15	21,60
GS309	11	16,17	GS347	11,17	10
GS310	17,27	21,22			
GS311	35,37	17,28			
GS312	11,22	10,21			
GS313	11	10			
GS314	14,27	10,15			
GS315	15,17	10			
GS316	11,25	20			
GS317	11,15	20,26			
GS318	11,26	10,21			
GS319	11	21			
GS320	11	16,21			
GS321	11,25	10,61			
GS322	14,17	57,60			
GS323	17,83	10,16			
GS324	11,36	10,21			
GS325	11	20			
GS326	11,14	-			
GS327	11	10,20			
GS328	11,17	10,20			
GS329	15,82	10,57			
GS330	11	21			
GS331	11,24	10			
GS332	15,24	20,28			
GS333	11	10,20			
GS334	11,15	17,20			
GS335	17	10			
GS336	11,17	21,52			
GS337	15	20,50			
GS338	14,17	58,60			
GS339	15,17	16			
GS340	11,14	10,21			
GS341	17,23	10,64			
GS342	14	21			

Appendix Two

DM-1 Single Sperm data

Allele	Av ALC	Allele	Av MR
5	0	5	0
6		6	
7		7	
8	0	8	0
9		9	
10		10	
11		11	
12	0	12	0
13		13	
14	0	14	0
15		15	
16		16	
17		17	
18		18	
19		19	
20		20	
21	0	21	0
22	0	22	0
23	0	23	0
24	-0.089	24	7.6
25	-0.0395	25	3.93
26	-0.032	26	4.8
27	-0.021	27	16.185
28	-0.035	28	12.75
29	-0.1925	29	14.3
30	0.007	30	15.64
31		31	
32	-0.103	32	16.71
33		33	
34	0.3805	34	37.55
35	0.581	35	58.2
36		36	
37		37	
38		38	
39		39	
40		40	

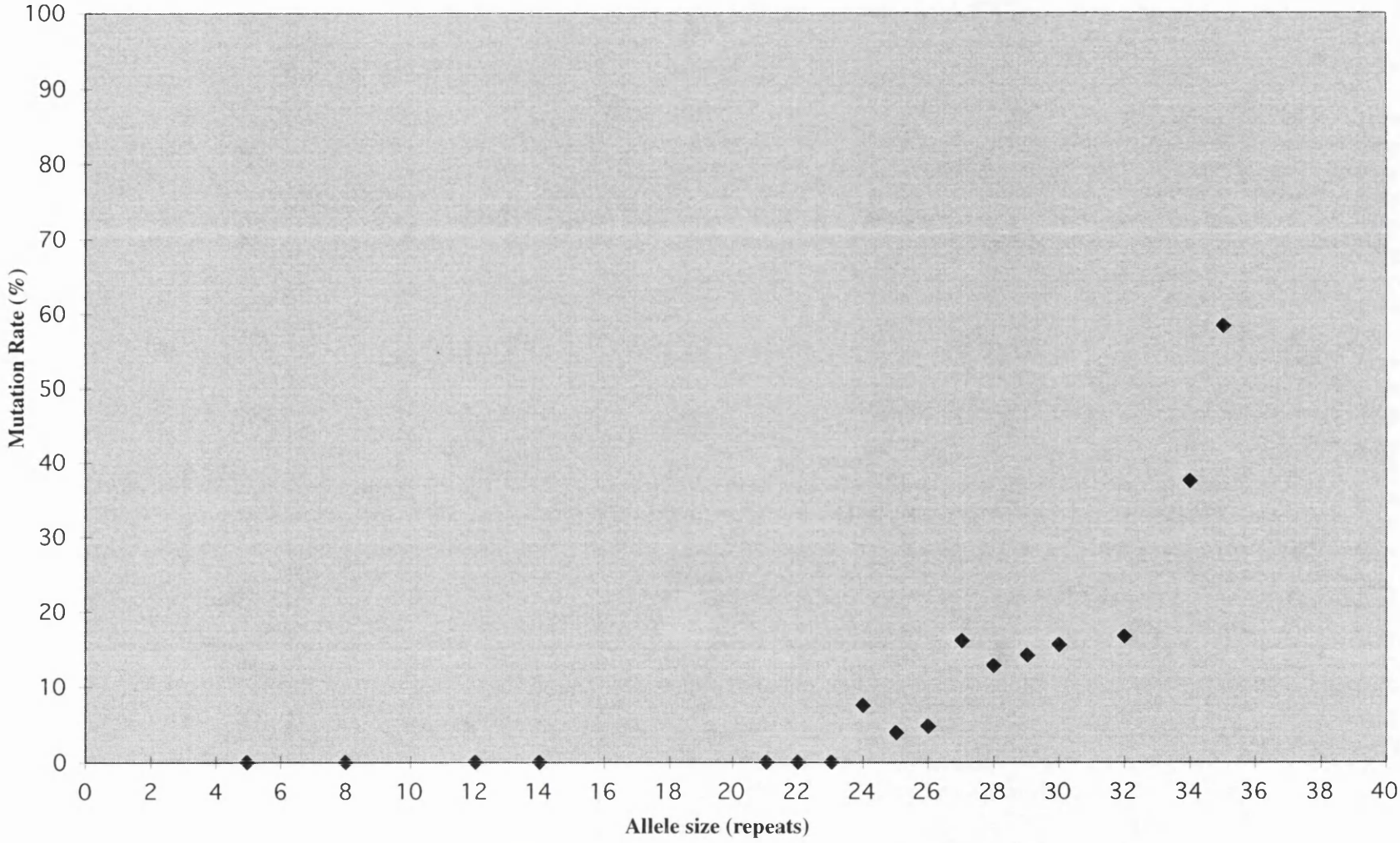

Figure A2.1 Average mutation rates of DM-1 alleles in the male germline.

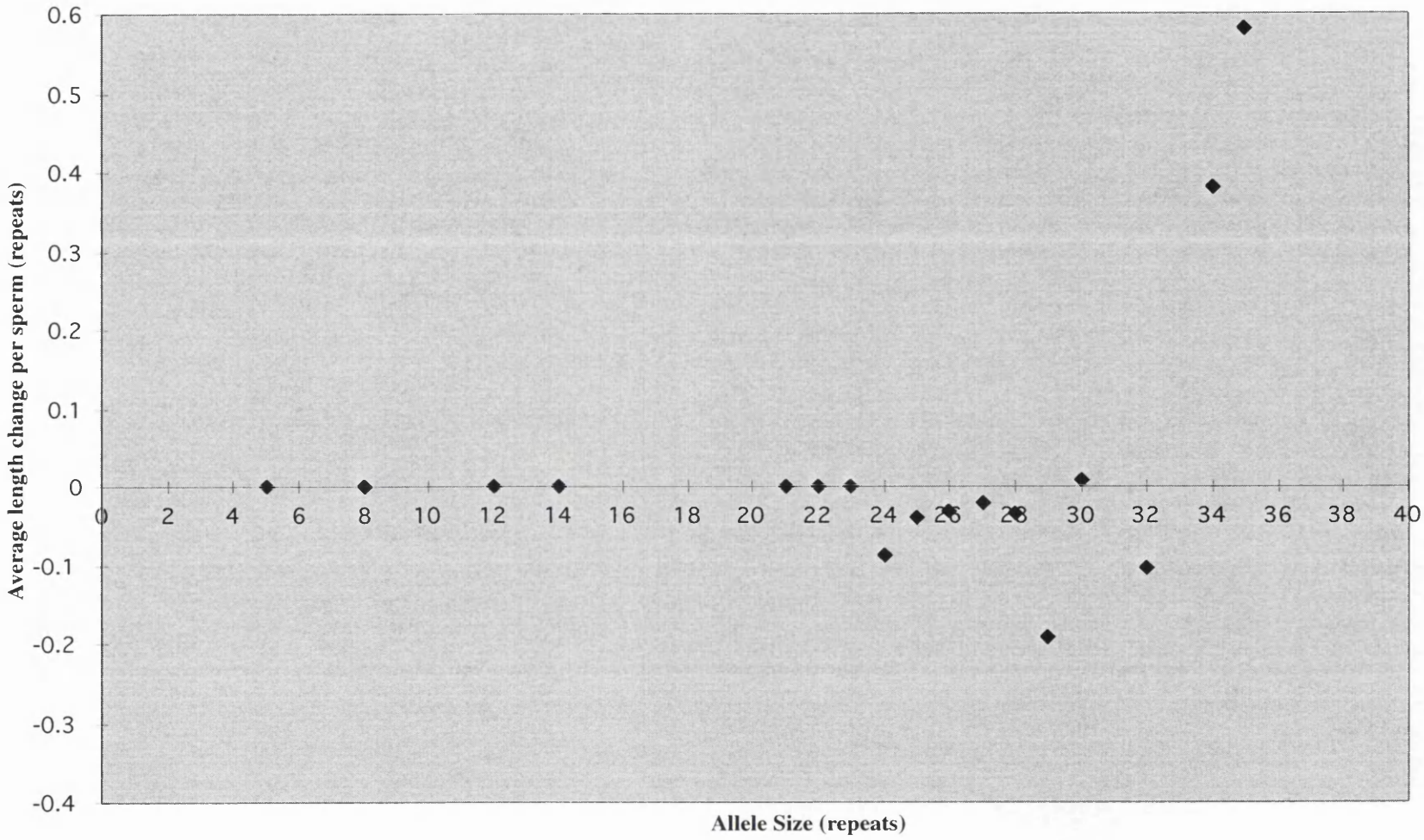

Figure A2.2 Average length change of DM-1 alleles in the male germline.