THE ROLE OF TELOMERASE AND CHROMOSOME HEALING IN PATIENTS WITH TERMINAL DELETIONS

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by
Helen Varley
Department of Genetics
University of Leicester

July 1999

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

Do or do not; there is no try.

Yoda. A long time ago...

CONTENTS

Chapter 2 MATERIALS AND METHODS

Chapter 3 ANALYSIS OF ORDERED ARRAY LIBRARIES GENERATED FROM PATIENTS WITH SUSPECTED TERMINAL DELETIONS

Chapter 4 INVESTIGATION OF THE TELOMERE-ANCHORED PCR AND CLONING STEPS USED TO CONSTRUCT ORDERED ARRAY LIBRARIES

Chapter 5 A FILTER HYBRIDISATION SELECTION STEP TO ENRICH FOR TELOMERE REPEAT ARRAYS IN ORDERED ARRAY LIBRARIES

Chapter 6 ANALYSIS OF SEQUENCE AROUND TERMINAL DELETION BREAKPOINTS AT 7q32, 22q13.3 AND 16p13.3

Chapter 7 FINAL DISCUSSION AND FUTURE WORK

THE ROLE OF TELOMERASE AND CHROMOSOME HEALING IN PATIENTS WITH TERMINAL DELETIONS

Helen Varley

Abstract

A simple telomere-anchored PCR-based method was used to isolate telomere-junction clones from a number of individuals. Analysis of clones led to the isolation of novel chromosome ends, including a low level polymorphic telomere, present in 2.4 % of the Caucasian population. The sequence adjacent to this telomere shows homology to a subterminal repeat found at internal locations on 4p, 4q, 10q and 22q in most individuals. A number of modifications were made to the telomere-anchored PCR strategy, including generation of telomere variant repeat probes to detect many normal chromosome ends, and the addition of a filter hybridisation selection step to increase enrichment for telomere-repeat arrays from genomic DNA. The modified strategy was used successfully to isolate a candidate terminal deletion breakpoint from a patient with a deletion of 22q. The telomere-adjacent sequence does not show homology to subterminal repeats and is unique to chromosome 22. It shows 96.2 % sequence identity to a BAC clone mapped to 22q13.3, consistent with initial RFLP data obtained by another group. The sequence is located adjacent to a telomere in the patient, but not in 87 unrelated individuals, or the patient's parents, indicating that it is a *de novo* telomere repeat array. The telomere appears to contain TTAGGG repeats, without variants. The lack of subterminal sequence at the breakpoint, and array of only TTAGGG repeats strongly suggests that the truncated chromosome was healed via direct addition of telomere repeats by the enzyme telomerase. Comparison of the sequence around this putative breakpoint region with sequence generated from a 7q32 breakpoint, and previously characterised breakpoints at 16p13.3, did not reveal any sequence features common to these regions, apart from a high density of dispersed repeats. The lack of conserved sequence features suggests that telomerase does not have strict sequence requirements for healing broken chromosomes *in vivo*.

Acknowledgements

Thanks go to all members of lab G19, including Alec, Angie, Carole, Caz, Celia, Dave, Duncan, Esther, Gemma, Ila, Jérôme, Jo C, John M, John S, Keiji, Maria, Philippe, Rita, Ruth, Yuri (the stats guru) for much appreciated help and advice, particularly the girls in the bay: Helen P, Hilda and Jenny F. Thanks also to everyone in the Department of Genetics that keep it running smoothly, and the inspirational standard of work that goes on in G19 in a relaxed and approachable atmosphere.

Extra thanks go to Nicola for being an excellent supervisor, guidance whenever needed, and getting me to the end of my PhD.

On a social note, cheers to the girls and boys for those Friday nights 'just for one' down the New Road Inn, the usual suspects including many of the respectable scientists listed above. A big thank-you goes out to the girls for many good times and much hilarity, especially Wendy (top housemate, our rantings and cups of tea), Emma (a repetitive beats comrade), Ruth and Lynn (marvellous nosh), Caz (being in the same boat), Liz and Jenny (those stories) and Philippe (honorary girl), and all of them for beers, gossip, and silly nights out.

On a personal note, extra special thanks to Mum, Dad and Kathryn for support and encouragement all the way, the lasses in Nottingham, Claire, Kerry and Cath for many a classy night out and enduring friendship, and the boys in Manchester for some rockin' weekends. Much appreciation goes to Nick (the retro kid) for his theories, and being a damn good mate when I'm not, and vice versa. Most impressive thanks go to Dave (lovely pair of decks), for letting me get on with it, keeping me sane at weekends, much love, and distracting me in all the right places.

PUBLICATIONS

Some of this work has been published previously.

Varley, H., Royle, N. J., Coleman, J. (1998)
The role of telomerase and chromosome healing in patients with terminal deletions.
Am. J. Hum. Gen. **s63**, A869

Abbreviations

A	adenine
ATP	adenosine 5'-triphosphate
ATR-16	alpha-thalassaemia and mental retardation syndrome
BAC	bacterial artificial chromosome
bp, kb, Mb	base pairs, kilobases, megabases
BSA	bovine serum albumin
C	cytosine
CEPH	Centre d'Etude du Polymorphism Humain
cm	centimetre
dATP	2'-deoxyadenosine 5'triphosphate
dCTP	2'-deoxycytidine 5'triphosphate
dGTP	2'-deoxyguanosine 5'triphosphate
dTTP	2'-deoxythymidine 5'triphosphate
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetra-acetic acid
EST	expressed sequence tag
FISH	fluorescence *in situ* hybridisation
G	guanine
g, mg, µg, ng	grams, micrograms, nanograms
HCl	hydrochloric acid
HPE	holoprosencephaly
hTR/mTR	human/mouse telomerase RNA component
hTRF	human telomere repeat binding factor
hTRT	human telomerase catalytic component
l, ml, µl	litre, millilitre, microlitre
LTR	long terminal repeat
M, mM, µM	molar, millimolar, micromolar
min	minute(s)
mRNA	messenger RNA
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PD	population doubling
PEG	polyethylene glycol
PEV	position effect variegation
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
sec	second
SDS	sodium dodecyl sulphate
SSC	saline sodium citrate
STS	sequence tagged site
SV40	simian virus 40
T	thymine
TAE	tris-acetate/EDTA buffer
TART	telomere-associated retrotransposon
TBE	tris-borate/EDTA buffer

TE	tris/EDTA buffer
TPE	telomere position effect
TRAP	telomere Repeat Amplification Protocol
TRF	telomere restriction fragment
Tris	tris-(hydroxymethyl)-methylamine[2-amino-(2-hydroxymethyl)-propan-1,3-diol]
TVR	telomere Variant Repeat
UV	ultraviolet
V	volts
x-gal	5-bromo-4-chloro-3-inolyl-(β)-D-galactopyranose
YAC	yeast artificial chromosome

Chapter 1

INTRODUCTION

Telomeres

Telomeres were first described by Herman Muller during a series of experiments using X-irradiated *Drosophila melanogaster* cells. Muller observed that ring chromosome formation involved loss of a small amount of terminal chromatin, allowing the chromosomes to stick together. Broken chromosomes were lost even if they had a centromere because their ends did not consist of natural termini. Muller hypothesised that the terminal DNA must have a special function of sealing the end of the chromosome - a region he called the telomere (Muller, 1938). Barbara McClintock also observed that in maize (*Zea mays*), broken chromosomes appeared to be unstable, adhesive at the ends, and tended to fuse with each other (McClintock, 1941). McClintock hypothesised that these breakages occurred during the normal process of crossing-over in meiosis, and that the breaks would eventually heal via acquisition of a telomere, thus discontinuing the breakage-fusion-bridge cycle that broken chromosomes cycled through. This pioneering work on telomeres demonstrated that the behaviour of natural ends of chromosomes was different to that of broken chromosome ends, and that broken ends without a telomere are generally unstable in a cell, subject to degradation or illegitimate recombination (reviewed Blackburn, 1984; Zakian, 1989).

The molecular structure of telomeres was first studied in the unicellular ciliate *Tetrahymena*, and described as an unusual repeated sequence motif present at chromosomal termini (Blackburn and Gall, 1978). Human telomeres were identified some years later, with the isolation of a highly conserved repetitive DNA sequence from a human recombinant DNA library found to be present on all human chromosomes, with major clusters at the termini. Human telomeres showed similarity to functional telomeres isolated from lower eukaryotes. Telomeric sequences were found to be highly conserved, from unicellular organisms to vertebrates (Moyzis *et al.*, 1988). Evolutionary conservation of human telomeric repeats was traced back in more than 100 species to a common ancestor that may have existed 400 million years ago (Meyne *et al.*, 1989).

The structure of telomeres

Telomeres consist of simple tandem repeat arrays of T_2AG_3 in humans (Moyzis *et al.*, 1988), chimpanzee (Luke and Verma, 1993), many mammals (de la Sena *et al.*, 1995; Meyne *et al.*, 1989), other vertebrates, and some invertebrates (Coleman *et al.*, 1993). A telomeric probe containing both the *Schizosaccharomyces pombe* ($TTAC(A)G_{1-6}$) and *Tetrahymena thermophila* (T_2G_4) repeats hybridised to a series of discrete fragments in DNA from a wide range of eukaryotes (Allshire *et al.*, 1988). The T_2AG_2 telomeric repeat is widely conserved among insects (Okazaki *et al.*, 1993). Table 1.1 shows telomere repeat sequences found in a variety of organisms. As more telomeric sequences become known it is increasingly difficult to identify even a loose consensus sequence to describe them. However, there appears to be a conserved strand bias running 5′ to 3′ with more G residues, which tend to be clustered. Telomeres generally have short repeat units arranged in a tandem array (Zakian, 1995).

Single cell eukaryotes. Telomeres were initially studied at the molecular level in ciliates, because single macronuclear cells contain from 40,000 to 100,000 telomeres. The high copy number of terminal sequences in these single cell eukaryotes made study much easier than in vertebrate cells, which often contain less than 100 chromosome ends per cell (reviewed Greider, 1990). In ciliate macronuclei such as *Tetrahymena*, DNA molecules are linear and require telomeric structure at the termini to be maintained and replicated. New telomeres are generated every time a new macronucleus develops (Prescott, 1994), and each of these DNA molecules has between 20 and 70 tandem repeats of T_2G_4 at the terminus (Blackburn *et al.*, 1983). Micronuclear telomeres contain the same tandem repeats (T_2G_4) as macronuclear telomeres but are approximately seven times longer (Kirk and Blackburn, 1995). In ciliates *Oxytricha*, *Euplotes* and *Tetrahymena,* the G-rich strand is oriented 5′ to 3′ towards the terminus, and the 3′ end protrudes 12-16 nucleotides beyond the C-rich strand to form a single-stranded tail (Henderson and Blackburn, 1989; Klobutcher *et al.*, 1981). These G-tails exist during most of the cell cycle and may be important for control of telomere length by the telomere maintenance machinery, which requires an exposed 3′ terminus (Lingner and Cech, 1996). They may also protect ends from fusion and degradation, and are required for binding of some telomere-associated proteins.

Yeast. *Saccharomyces cerevisiae* telomeres acquire transient TG_{1-3} tails in late S phase. These tails may exist during the rest of the cell cycle at lengths of less than 30 nucleotides (Wellinger *et al.*, 1992; 1993; 1996). The loss of telomeric sequence

Table 1.1 Telomere repeat sequences found in a variety of organisms

Organism	Repeats	Reference
Mammals		
Homo sapiens	TTAGGG	(Moyzis *et al.*, 1988)
Mus musculus	TTAGGG	(Meyne *et al.*, 1989)
Invertebrates		
Ascaris lumbricoides	TTAGGC	(Muller *et al.*, 1991)
Plants		
Arabidopsis thaliana	TTTAGGG	(Richards and Ausubel, 1988)
Algae		
Chlamydomonas reinhardtii	TTTTAGGG	(Petracek *et al.*, 1990)
Funghi		
Saccharomyces cerevisiae	TG_{1-3}	(Shampay *et al.*, 1984)
Schizosaccharomyces	$TTAC(A)G_{2-5}$	(Matsumoto *et al.*, 1987)
Protozoa		
Oxytricha	TTTTGGGG	(Klobutcher *et al.*, 1981)
Euplotes	TTTTGGGG	(Klobutcher *et al.*, 1981)
Paracmecium	TTT(T/G)GGG	(Forney and Blackburn, 1988)
Tetrahymena	TTGGGG	(Blackburn and Gall, 1978)
Trypanosoma	TTAGGG	(Van der Ploeg *et al.*, 1984)
Slime Mould		
Dictyostelium	AG_{1-8}	(Emery and Weiner, 1981)
Physarum	TTAGGG	(Bergold *et al.*, 1983)
Insects		
Bombyx mori (silkworm)	TTAGG	(Okazaki *et al.*, 1993)

per *S. cerevisiae* cell division is consistent with a model in which both ends of yeast chromosomes have an approximately 10 nucleotide G-rich overhang (Zakian, 1996).

Humans. Human telomeres do not only contain uniform six base pair repeats. There are many types of repeat present at the proximal ends of human chromosomes including T_2G_4 and T_3AG_3 (Allshire *et al.*, 1989), $TCAG_3$ and $TGAG_3$ (Guerrini *et al.*, 1993). Distribution of these repeats appears to be non-random, and may result from mutation processes taking place at human telomeres (Allshire *et al.*, 1989). Variant repeats tend to be clustered at the proximal end of the telomere, possibly because this region would be less likely to be exposed and corrected by the telomere maintenance machinery (Allshire *et al.*, 1989; Guerrini *et al.*, 1993). Existence of these variant repeats at the distal end of the telomere may be limited by telomere structure, or telomere binding proteins that require $(T_2AG_3)n$ repeats for efficient binding, discussed later in this chapter. A high proportion (80-90 %) of human chromosome ends also have G-rich overhangs (Makarov *et al.*, 1997), on average 130-210 bases in length. Arrays of $(T_2AG_3)n$ are also present at numerous interstitial sites in the genome (Wells *et al.*, 1990). These telomere-like sequences share nucleotide sequence similarity outside the repetitive array, suggesting evolution from a common progenitor locus. There is an interstitial array of degenerate TTAGGG-like repeats at 2q13-q14, where the fusion of ancestral ape chromosomes is thought to have occurred (IJdo *et al.*, 1991). In the ciliate *Paramecium*, interstitial telomere-like sites appear to be hotspots for illegitimate recombination with injected foreign DNA molecules (Katinka and Bourgain, 1992).

The chromatin structure of telomeres

Saccharomyces cerevisiae telomere repeat tracts (Wright *et al.*, 1992) and ciliate macronuclear DNA molecules (Blackburn and Chiou, 1981; Gottschling and Cech, 1984) are packed in a discrete, non-nucleosomal chromatin structure called the telosome. In mammals, most of the telomeric DNA is packaged in nucleosomes more tightly packed than bulk nucleosomes (Makarov *et al.*, 1993). However, the very ends are thought to be packaged in telosome-like structures, and human telomeres display unusual chromatin structure consistent with a telosome-like configuration when they are relatively short (2-7 kb). Regions adjacent to the telomere appear to have normal nucleosome organisation (Tommerup *et al.*, 1994).

Physical studies show that telomeric DNA oligonucleotides can form non-Watson-Crick G-G base pairs and self associate, behaving like hairpin duplexes, leading to unusual 'G-quartet' intramolecular structures with hydrogen bonding (Henderson *et al.*, 1987; Sundquist and Klug, 1989). This 'G-quartet' structure is

readily formed under physiological conditions and may be important for telomere functions such as telomere-telomere association and recombination (Williamson *et al.*, 1989).

Telomere position effect

In *S. cerevisiae* (Gottschling *et al.*, 1990) and *S. pombe* (Nimmo *et al.*, 1994) the expression of genes placed adjacent to a telomere can be reversibly repressed. This is known as the telomere position effect (TPE). The extent of transcriptional silencing is dependent upon the strength of the gene promoter, distance from the telomere and length of the telomere (Gottschling *et al.*, 1990; Kyrion *et al.*, 1993; Renauld *et al.*, 1993). This position dependent gene expression is also known as position effect variegation (PEV), and was first identified in *Drosophila*. It is usually the result of a genomic rearrangement relocating a gene close to a region of heterochromatin and causing gene repression, for example, the *Drosophila* white gene mutation affects eye colour (reviewed Henikoff, 1992). In *S. cerevisiae*, histone H4 mutations destroy the telomere position effect, directly implicating chromatin structure in the silencing mechanism (Gottschling, 1992). TPE also involves a number of proteins including the transcriptional silencing proteins SIR2, SIR3 and SIR4, and mutations in these genes remove telomere position effect (Aparicio *et al.*, 1991). Internal tracts of telomeric sequence can also cause transcriptional repression (Zakian, 1995).

Fascioscapulohumeral dystrophy (FSHD) is an autosomal dominant neuromuscular disorder in humans, linked to the most distal genetic markers on 4q35. An *Eco*RI fragment length polymorphism segregates with the disease in most families and a 3.3 kb tandem repeat array (D4Z4) lies in this fragment (Wijmenga *et al.*, 1992), within 5-14 kb of the telomere (Bengtsson *et al.*, 1994). Deletions of more than 100 kb in this repeat are associated with FSHD (van Deutekom *et al.*, 1993). It was hypothesised that deletions in this repeat caused transcriptional silencing (PEV) in this region by moving the telomere to a more proximal location, or by disruption of the usually heterochromatic structure of the region. However, no evidence for transcription repression of the FRG1 gene (FSHD region gene 1), located 100 kb proximal to the repeat, has been found (van Deutekom *et al.*, 1996). It is not clear whether human telomeres exert a position effect (Bayne *et al.*, 1994). Expression of gene constructs in human cell lines inserted 4kb from the telomere were unaffected by telomeres of length ranging from 0.5 to 25 kb, suggesting that chromatin differences at telomeres do not confer TPE in mammals (Sprung *et al.*, 1996).

Telomere Function

1. Chromosomal stability and protection

Telomeres are essential for chromosome stability and protective 'capping' of the chromosome terminus (McClintock, 1941; Muller, 1938). The role of telomeres in chromosome stability has been demonstrated *in vivo* and *in vitro* in *Xenopus laevis* eggs (Weber *et al.*, 1993) and *Paramecium* macronuclei (Bourgain and Katinka, 1991). Linear DNA constructs capped by human telomeres are stable, in contrast to constructs with non-telomeric ends, which undergo intramolecular fusion and degradation (Bourgain and Katinka, 1991; Li *et al.*, 1998). In wild type yeast cells, elimination of a telomere caused cell cycle arrest, demonstrating that telomeres help cells distinguish intact chromosome ends from broken or damaged chromosomes (Sandell and Zakian, 1993).

2. Nuclear organisation

Telomeres are not randomly distributed within the nucleus, and spatial distribution of mammalian telomeres appears to vary according to cell type and cell cycle stage (Manuelidis and Borden, 1988). During mitotic interphase human telomeres are distributed throughout the nuclear volume (Manuelidis and Borden, 1988; Vourc'h *et al.*, 1993), but are tightly associated with the nuclear matrix (which includes the nuclear envelope) by the $(T_2AG_3)n$ repeats (de Lange, 1992). Telomeres participate in mitotic chromosome segregation (Rasmussen and Holm, 1980) and telomere defects can impair this process (Kirk *et al.*, 1997). Human telomeres redistribute during the cell cycle, with distal telomeres (chromosome long arms) moving from the interior of the nucleus to the periphery between G1 and G2 phases, and proximal telomeres and centromeres moving in the opposite direction (Vourc'h *et al.*, 1993). In contrast, telomere clustering near the nuclear membrane has been observed in many animal species during first meiotic prophase, forming the 'bouquet' structure. Meiotic pairing is often initiated from telomeric regions (Rasmussen and Holm, 1980), consequently telomere-telomere interactions and attachments to the nuclear envelope during meiosis may assist chromosome alignment and pairing (reviewed Loidl, 1990).

3. The end replication problem

The replication problem for linear DNA exists because all known DNA polymerases are unidirectional (synthesis occurs in the 5′ to 3′ direction) and require a 3′-OH group to add nucleotides to a DNA strand (Watson, 1972; Olovnikov, 1973; Zakian, 1995). For normal lagging strand replication, short RNA primers provide the 3′-OH group for DNA polymerase, which extends them to generate Okazaki fragments. The RNA primers are removed from the lagging strand when synthesis has been

carried out, leaving a 5′ gap at the terminus (reviewed Blackburn, 1984) at least the length of the RNA primer (8-14 bases). As a consequence, loss of sequence from the ends of chromosomes occurs during each round of DNA replication (Olovnikov, 1971; Olovnikov, 1973; Watson, 1972). The end replication problem is shown in figure 1.1. Telomeres act as a 'buffer' to compensate for this sequence loss at chromosome ends, protecting adjacent DNA which may contain essential sequence.

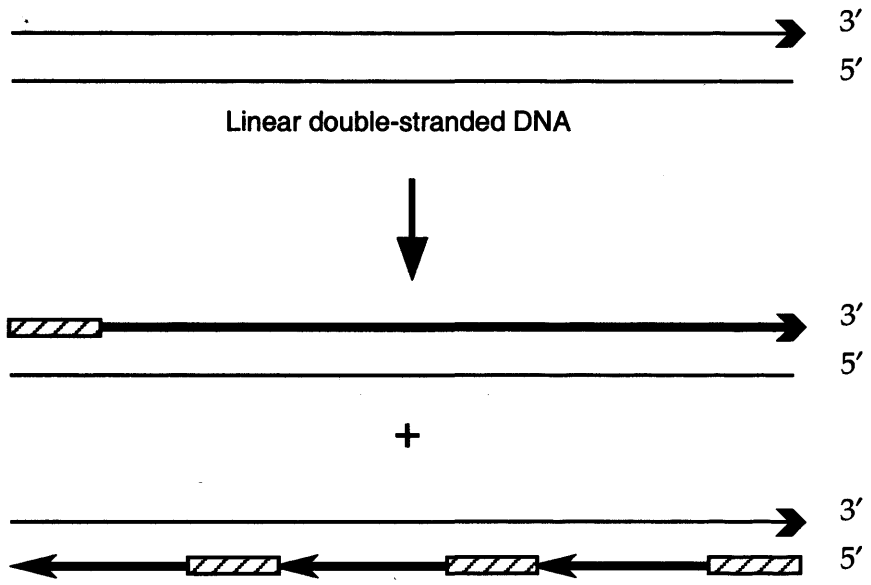
The unusual termini of *Drosophila* chromosomes

Isolation of the terminal DNA of *Drosophila* chromosomes established that the termini lack short G-rich tandem repeats common to telomeres in other species, yet have the same functional properties as other eukaryotic telomeres, and prevent loss of essential sequence during chromosome replication. The terminal 14.5 kb of *Drosophila* chromosomes are composed of tandem elements derived from two families of non-Long Terminal Repeat (non-LTR) retrotransposons, and are subject to slow terminal loss. One of these elements is the Het-A family of transposable elements, which have exclusive telomeric location (Pardue *et al.*, 1996). Approximately 1 % of chromosome ends receive a new Het-A element each generation. The estimated minimal size of these elements is 6 kb, so approximately 60 bp are added to each chromosome per generation. This is enough to balance the progressive loss of 70-80 bp per generation due to incomplete end replication (Biessmann *et al.*, 1992; reviewed Mason and Biessmann, 1995). TART (telomere-associated retrotransposon) elements are also found in *Drosophila* (Levis *et al.*, 1993; reviewed Pardue *et al.*, 1996) and transpose preferentially to chromosome ends as part of the process of telomere maintenance (Sheen and Levis, 1994).

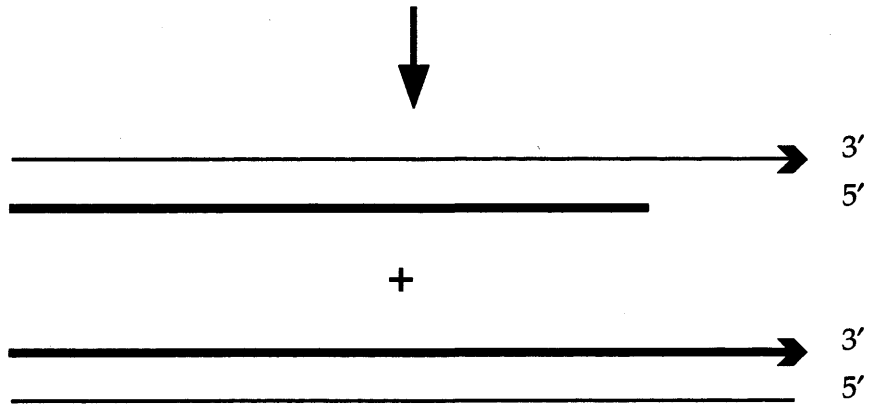
Telomere dynamics

Telomere length is very variable across species. Human telomeres in somatic cells range from 2-15 kb (Counter *et al.*, 1992; Hastie *et al.*, 1990; Moyzis *et al.*, 1988), and telomeres in sperm up to 20 kb in length have been detected (Cross *et al.*, 1989; de Lange *et al.*, 1990; Hastie *et al.*, 1990; Lejnine *et al.*, 1995). In the ciliate *Oxytricha*, telomeres are only 20 bp long (Klobutcher *et al.*, 1981) and in yeast, telomere tracts range from 200-400 bp (Shampay *et al.*, 1984). Telomere length is estimated by digesting genomic DNA with a restriction enzyme which does not cut within the telomeric repeat array, such as *Hae*III or *Hinf*I, then blotting using the Southern method (Southern, 1975) and probing with a telomere repeat probe, giving a characteristic smear of hybridising products called Telomere Restriction Fragments (TRFs). A considerable variation in TRF length has been observed

Figure 1.1 The end replication problem



Semi-conservative DNA replication. New DNA synthesis is primed by short RNA primers. Leading strand synthesis results in complete replication of the upper strand. Lagging strand synthesis results in generation of Okazaki fragments.



Removal of RNA primers and ligation of Okazaki fragments to fill gaps. Lagging strand 5' end cannot be synthsised.

among unrelated individuals (Slagboom and Vijg, 1992). Mean TRF length from human monozygotic and dizygotic twin pairs aged 2-95 years gave a heritability of 78 % for telomere length, indicating that individual differences in mean TRF length appear to be largely genetically determined (Slagboom *et al.*, 1994). Telomere fluorescence intensity values in normal adult bone marrow showed that the telomere signal from sister chromatids appeared to be of similar intensity (Lansdorp *et al.*, 1996).

Due to the end replication problem, telomeric sequence is lost with every cell division. This was detected by a decrease in the length of telomere repeat arrays of somatic cells with an increase in the age of the donor (Hastie *et al.*, 1990). The mean rate of telomere reduction was calculated to be approximately 33 bp per year in normal blood and colon mucosa (Hastie *et al.*, 1990; Slagboom *et al.*, 1994) and approximately 41 bp in peripheral blood lymphocytes (Vaziri *et al.*, 1993). TRF length in most human fetal tissues was found to be similar, although there were significant variations among fetuses. This synchrony in telomere length is lost after birth (Youngren *et al.*, 1998) as telomeric sequence is lost with each cell division. Comparing human fetal liver and adult bone marrow primitive progenitor cells showed a highly significant difference in the mean TRF length between fetal and adult tissues (Vaziri *et al.*, 1994). Human cells in culture also lose telomeric DNA as a function of the number of population doublings (PDs). Mean TRF length was shown to decrease in cultured normal lymphocytes at a rate of 85-155 bp per PD (Vaziri *et al.*, 1993). In fibroblasts, estimated rates of telomere loss *in vivo* (~75bp per PD) and *in vitro* (28-70bp per PD) are comparable (Harley *et al.*, 1990; Allsopp *et al.*, 1992). Others factors such as enzymatic degradation may also contribute to the erosion of telomere length, as mammalian cells lose telomeric DNA about 10 times faster than expected from incomplete replication alone (Zakian, 1997). However, cells that divide repeatedly, for example malignant or neoplastic cells and germline cells, restore their telomeres by adding telomere repeats at every cell division (reviewed Kipling, 1995).

Mus spretus have a mean TRF length similar to that of humans, ranging from 5-20 kb (Starling *et al.*, 1990). However, *Mus musculus* telomeres are estimated to be 8-16 times longer than human repeat arrays. They are resolved as multiple discrete restriction fragments of up to 150 kb and do not perceptibly shorten during the lifespan (Kipling and Cooke, 1990; Starling *et al.*, 1990). Telomere length was found to be similar among different tissues within a newborn mouse but differed between tissues in an adult mouse. Telomere length also decreased with increased doublings in mouse cell cultures, which correlates with human TRF data (Prowse and Greider, 1995).

Telomere binding proteins

Single cell eukaryotes. It is proposed that components of the telomeric chromatin wrap the telomere up into a structured complex preventing unlimited access of the telomere-maintenance machinery to the telomere (Lundblad and Wright, 1996). A number of proteins are associated with telomeric sequences *in vivo* and are thought to be essential for function. In the ciliate *Oxytricha nova*, structural proteins recognise and bind tightly (but not covalently) to the G-rich single strand overhang sequence forming a protective complex at the very end of the telomere (Zakian, 1995). One example is the α/β heterodimer TEBP (telomere end binding protein; Gottschling and Zakian, 1986). The crystal structure of TEBP has been determined, and the α and β units form a deep groove. This groove folds the single strand telomere 3′ overhang to form a loop that is completely buried within the complex, so that it is inaccessible to the telomere maintenance machinery *in vivo* (Horvath *et al.*, 1998). Similar proteins have been isolated in *Euplotes crassus* (Price, 1990), *Chlamydomonas reinhardii* (Petracek *et al.*, 1994) and *Tetrahymena thermophila* (Sheng *et al.*, 1995). The *Xenopus* telomere end factor protein (XTEF) specifically recognises vertebrate telomere repeats at DNA ends when $(T_2AG_3)_2$ is present in single stranded form, and its binding properties resemble the terminus specific telomere proteins seen in ciliates (Cardenas *et al.*, 1993). However, similar proteins have not yet been isolated in humans.

Yeast. Yeast have some telomere binding proteins with different protein-DNA binding properties to those found in ciliates (Wright and Zakian, 1995). TBF-α is an abundant protein, essential for growth in yeast, and binds at the junction of subtelomeric sequence and $(TG_{1\text{-}3})$ repeats (Liu and Tye, 1991). However, TBF-α deletion mutants with severe growth defects fail to show any detectable change in telomere length (Brigati *et al.*, 1993), suggesting that it may not be associated with telomere function *in vivo*. This protein also binds to telomeric sequences from *Tetrahymena* and vertebrates (Liu and Tye, 1991).

Two other proteins associated with telomeres in *S. cerevisiae*, are encoded by the TEL1 and TEL2 genes. Cells with tel-1 mutations had short telomeres (Greenwell *et al.*, 1995; Runge and Zakian, 1996), and did not respond to changes that caused telomere lengthening in wild-type cells. These cells also showed slightly elevated levels of chromosome loss and mitotic recombination (Greenwell *et al.*, 1995). TEL2 is an essential gene, and cells with tel-2 mutations also have short telomeres. The tel-2 mutation reduces PEV, suggesting that this protein is involved in assembling the chromatin structure of the telomere and adjacent DNA (Runge and Zakian, 1996).

The Rap1p (Repressor Activator Protein) protein co-precipitates with telomeric chromatin in *S. cerevisiae* (Conrad *et al.*, 1990). Rap1p is involved in transcriptional silencing (Kurtz and Shore, 1991) and the RAP1 gene is essential for cell viability (Shore and Nasmyth, 1987), but alterations in Rap1p levels also affect telomere length. Most chromosome-bound Rap1p is localised to the ends of meiotic chromosomes and binds sequence specifically to the (TG_{1-3}) strand of telomeres (Gilson *et al.*, 1993). Rap1p contains a Myb-like DNA binding repeat and an amino-terminal acidic domain - the 'telobox' (Bilaud *et al.*, 1996). When Rap1p binds at its recognition consensus, it appears to untwist double stranded DNA which may promote non-duplex strand interactions at telomeres (Gilson *et al.*, 1994). A number of other proteins bind Rap1p at the telomere. The yeast silencing machinery encoded by the SIR2, SIR3 and SIR4 genes forms a complex mapping to the same sites along telomeric heterochromatin as RAP1 in wild type yeast cells (Palladino *et al.*, 1993). SIR3 is a structural component of yeast heterochromatin, repressing adjacent genes as it spreads along the chromosome (Renauld *et al.*, 1993) and telomeres shorten in sir3 and sir4 mutants (Palladino *et al.*, 1993). The *S. pombe* telomere protein Taz1p shares homology with the telobox DNA-binding domain of Rap1p. Mutation of the TAZ1 gene results in massive elongation of telomere-repeat tracts and complete loss of telomere position effect (Kyrion *et al.*, 1992, 1993).

Humans. The human telomere repeat binding factor (hTRF1) associates with double-stranded T_2AG_3 repeat arrays *in vitro* and is detected mainly at chromosome ends (Cardenas *et al.*, 1993) but does not require an end to bind. The protein and mRNA were detectable in all cells and tissues studied (Zhong *et al.*, 1992). TRF1 is an integral part of the telomere complex and probably plays a role in the compact configuration of telomeres (Luderus *et al.*, 1996). It binds preferentially to long contiguous repeats and the optimum site is more than six consecutive T_2AG_3 repeats, consistent with TRF1 binding along the length of mammalian telomeres *in vivo*. TRF1 shows a high degree of specificity, preferring T_2AG_3 repeats over variant repeats and closely related sequences (Zhong *et al.*, 1992). TRF1 contains a telobox DNA binding domain and binds in a dimer, distorting the telomeric DNA where it binds (Bianchi *et al.*, 1997), similar to Rap1p in *S. cerevisiae*. The TRF1 gene is involved in limiting telomere elongation by the telomere-maintenance enzyme telomerase. Long term over-expression of wild-type TRF1 in a tumour-cell line with stable telomeres resulted in a gradual and progressive telomere shortening. Telomere elongation was induced by expression of a dominant negative TRF1 mutant that inhibited binding of endogenous TRF1 to telomeres (van Steensel and de Lange, 1997). The TRF2 gene is a distant homologue of TRF1, with a similar telobox binding domain at the C-terminal. It is ubiquitously

expressed and also specifically binds to double stranded T_2AG_3 repeats *in vitro*, localising to all human telomeres on metaphase chromosomes. The telomere binding activity of TRF2 also shows high sequence specificity, binding more conservatively to T_2AG_3 than TRF1 (Broccoli *et al.*, 1997b). However, these proteins show significant differences and they have distinct functions (Broccoli *et al.*, 1997b). Part of the TRF2 sequence is related to the small non-histone chromatin factor HMG17, suggesting a role in the unusual nucleosomal organisation observed in telomeres (Bilaud *et al.*, 1997). Mouse TRF1 (mTRF1) was also isolated as a common component of all chromosome ends (Chong *et al.*, 1995) and is ubiquitously expressed in somatic and germline tissues (Bianchi *et al.*, 1997). hTRF1 and mTRF1 are similar sizes, with almost identical C-terminal Myb-like binding domains, but the rest of the proteins are poorly conserved suggesting rapid evolution (Broccoli *et al.*, 1997a).

The human telomere-associated protein Tankyrase (Smith *et al.*, 1998) shows homology to a class of proteins involved in protein-protein interactions. Tankyrase localises to human telomeres and binds TRF1 *in vitro*, causing a reduction in its ability to bind to telomeric DNA. Consequently, Tankyrase may prevent TRF1 from binding to telomeres so that telomerase has access to the chromosome end. Human telomere function may be regulated by this mechanism *in vivo* (Smith *et al.*, 1998). Other proteins have been implicated in telomere metabolism, such as Ku, involved in double-strand break repair (Boulton and Jackson, 1996) and the role these proteins may play in telomere function has yet to be elucidated.

Regulation of telomere length by Rap1p in *S. cerevisiae*

Telomere elongation is negatively regulated by Rap1p to keep telomere tracts within a given length distribution (Kyrion *et al.*, 1992), apparently by a simple protein counting mechanism (Marcand *et al.*, 1997). When hybrid proteins containing the Rap1p C-terminus were targeted to a telomere, the number of repeats was reduced in proportion to the number of targeted molecules, suggesting a feedback mechanism that can detect the exact number of Rap1p molecules binding to the telomere. Addition of Rap1p molecules on one side of a telomere resulted in a loss of the same number of molecules on the other side, indicating that the number of Rap1p molecules at a telomere may be maintained at a constant number. Rap1p may prevent telomerase binding or activity and when degradation or incomplete replication of a telomere causes loss of Rap1p binding sites, the telomere is elongated to add new Rap1p binding sites, and then returns to the repressed state (Marcand *et al.*, 1997). Sequences at the telomere junction affect length regulation by Rap1p, and it is thought that telomere length

regulation involves a folded chromatin structure which blocks access by telomerase and allows counting of Rap1p molecules (Ray and Runge, 1999). The counting mechanism appears to require other interacting proteins called Rif1 and Rif2 (Rap1-interacting factors) that bind to Rap1 and control telomere length regulation by the Rap1 protein. Deletion of either of the Rif proteins results in huge extension of telomeres (reviewed Barinaga, 1997) while over-expression decreases telomere length (Wotton and Shore, 1997).

It is likely that a similar protein counting mechanism is present in human cells. The Rap1p protein, the human TRF1 protein (Chong *et al.*, 1995) and the *S. pombe* Taz1p protein (Cooper *et al.*, 1997) all share similar telobox DNA-binding domains (Bilaud *et al.*, 1996), although outside these domains hTRf1 has very little homology to Taz1p, and no homology with Rap1p (Smith and de Lange, 1997; reviewed Shore, 1997). However, these proteins all bind to double stranded telomere repeat tracts, and TRF1 and Rap1p both distort DNA on binding, which may be required for correct telomere function in yeast and mammals (Bianchi *et al.*, 1997). Mutations in all three genes cause changes in telomere length (Conrad *et al.*, 1990; Kyrion *et al.*, 1992; 1993; van Steensel and de Lange, 1997). It is likely that all function similarly to negatively regulate telomere elongation by telomerase. The human Tankyrase protein may act in a similar way to the Rif proteins by regulating binding of TRF1 to the telomere (Smith *et al.*, 1998). The model for telomere length regulation by these telomere repeat binding proteins is shown in figure 1.2.

Telomere maintenance by telomerase

Single cell eukaryotes. Telomeres are extended and maintained by the enzyme telomerase, a ribonucleoprotein reverse transcriptase complex with RNA and protein components both essential for function. Telomerase synthesises the G-rich strand of the repeat array 5′ to 3′ (Greider and Blackburn, 1987, 1989; Morin, 1989), and contains a telomere-complementary sequence in the RNA component which is used as a template for adding repeats (Greider and Blackburn, 1987, 1989; Shippenlentz and Blackburn, 1990). Telomerase was first identified as a novel activity isolated in *Tetrahymena* cell free extracts, that added tandem T_2G_4 repeats onto synthetic telomere primers (Greider and Blackburn, 1985). Up to 8000 nucleotides of T_2G_4 repeats were added to the primer *in vitro* (Greider and Blackburn, 1989). A greater elongation activity was present during macronuclear development, when a large number of telomeres are formed and replicated, than during vegetative cell growth (Greider and Blackburn, 1985), indicating that

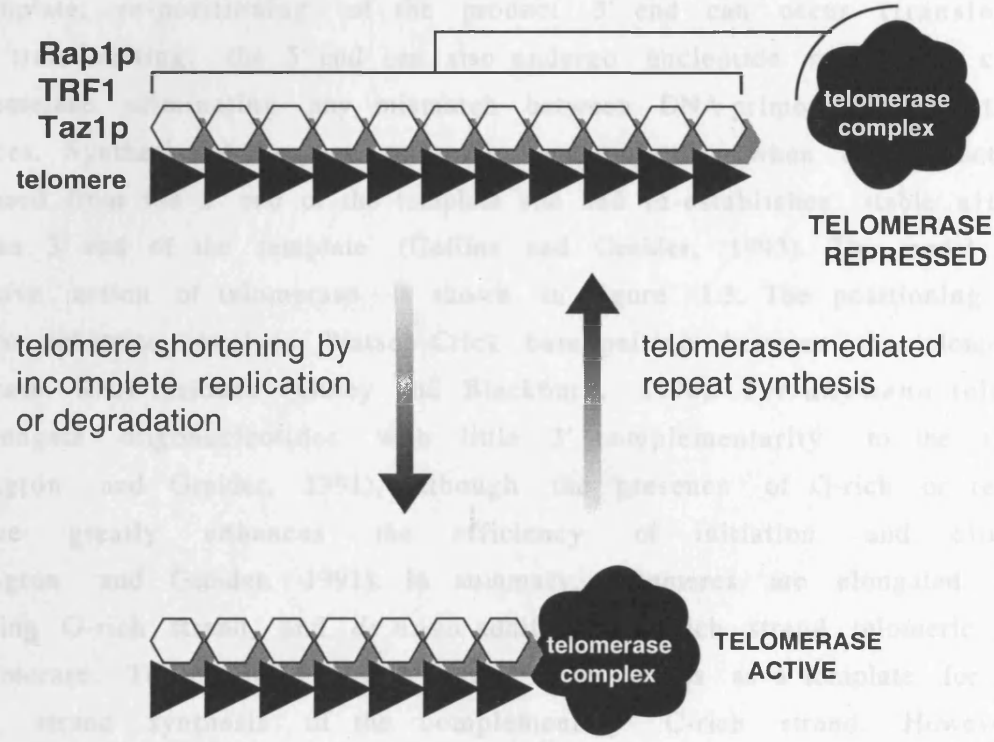


Figure 1.2 The model for telomere length regulation

Telomere length is thought to be regulated via double-stranded telomere repeat binding proteins (Rap1p/TRF1/ Taz1p) by a simple protein counting mechanism. A threshold number of binding protein molecules leads to repression of telomerase. Loss of telomere repeats causes a reduction in the number of bound protein molecules and telomerase repression is released, allowing synthesis of the telomere tract until the threshold number of protein binding molecules is reached. Telomere binding proteins Rap1p/TRF1/Taz1p are represented by white hexagons, and the shaded areas represent other binding proteins that are thought to interact such as the SIR and Rif proteins in *S. cerevisiae,* and Tankyrase in humans. Figure adapted from Shore (1997).

Tetrahymena telomerase adds repeats directly onto non-telomeric sequences during macronucleus development and controlled chromosome healing.

The model for telomere binding and elongation by *Tetrahymena* telomerase predicts that there are two distinct binding sites. One is the template site which binds at least four nucleotides at the 3′ end of the substrate sequence, and the second site (the anchor site) recognises telomere or telomere-like sequences and binds residues 5′ to the template site. After binding and elongation to the 5′ end of the template, re-positioning of the product 3′ end can occur (translocation). Before translocating, the 5′ end can also undergo nucleotide removal or cleavage by telomerase, eliminating any mismatch between DNA primer and RNA template sequences. Synthesis of another telomeric repeat begins when the product 3′ end is released from the 5′ end of the template site and re-establishes stable alignment with the 3′ end of the template (Collins and Greider, 1993). The model for the processive action of telomerase is shown in figure 1.3. The positioning of the telomere substrate involves Watson-Crick base pairing between the telomere and telomerase RNA residues (Gilley and Blackburn, 1996). *Tetrahymena* telomerase can elongate oligonucleotides with little 3′ complementarity to the template (Harrington and Greider, 1991), although the presence of G-rich or telomeric sequence greatly enhances the efficiency of initiation and elongation (Harrington and Greider, 1991). In summary, telomeres are elongated via the protruding G-rich strand, and *de novo* addition of G-rich strand telomeric repeats by telomerase. The extended G-rich strand then serves as a template for normal lagging strand synthesis of the complementary C-rich strand. However, the extension of the G-rich strand can only occur in cells that have telomerase activity (reviewed Blackburn, 1992b).

Humans. Human telomerase activity was first identified in crude HeLa cell extracts, and catalysed the addition of a six-nucleotide repeating pattern to oligonucleotide primers containing the human telomeric repeat sequence. In optimal conditions, 65-70 repeats were synthesised *in vitro* in a processive manner (Morin, 1989). Human telomerase can utilise oligonucleotides *in vitro* corresponding to the G-rich telomere repeats of *Oxytricha*, *Tetrahymena* and *Saccharomyces* for repeat addition with similar efficiency to (T_2AG_3) oligonucleotides (Prowse *et al.*, 1993), and some oligonucleotides with 3′ ends showing little complementarity to the template (Morin, 1991), suggesting that 3′ end pairing is not essential.

In 1994, a PCR-based method was developed to increase sensitivity and efficiency of detecting telomerase activity - the Telomere Repeat Amplification Protocol

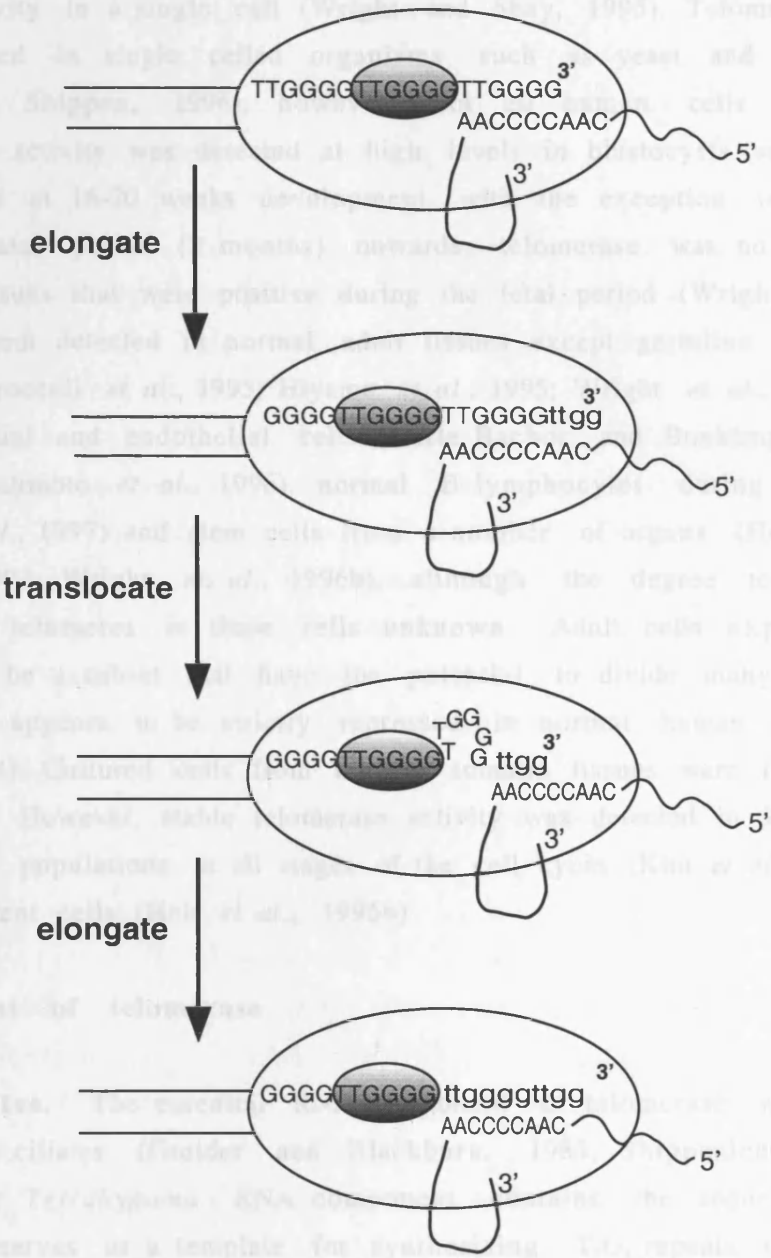


Figure 1.3 Model for processive telomerase action

The telomerase RNA template aligns against a substrate sequence and telomerase elongates the 3' terminus (lowercase letters). After the first round of elongation, when the extended product reaches the 5' end of the template, telomerase translocates along the substrate for a further round of elongation. The substrate is also bound at the 5' end by the 'anchor' site, indicated by a shaded oval. Figure adapted from Greider and Blackburn (1989) and Collins and Greider (1993).

(TRAP; Kim *et al.*, 1994). Telomerase synthesises extension products which then serve as templates for PCR amplification. TRAP is very sensitive - it can be used to detect telomerase activity in a single cell (Wright and Shay, 1995). Telomerase is constitutively expressed in single celled organisms such as yeast and ciliates (reviewed Melek and Shippen, 1996), however, not all human cells express telomerase. Telomerase activity was detected at high levels in blastocysts and most human somatic tissues at 16-20 weeks development with the exception of brain tissue. From the neonatal period (2 months) onwards, telomerase was no longer detected in somatic tissues that were positive during the fetal period (Wright *et al.*, 1996b). Telomerase is not detected in normal adult tissues except germline tissues, haemopoietic cells (Broccoli *et al.*, 1995; Hiyama *et al.*, 1995; Wright *et al.*, 1996b), skin epidermis, epithelial and endothelial cells (Harle-Bachor and Boukamp, 1996; Hsiao *et al.*, 1997; Yasumoto *et al.*, 1996), normal B lymphocytes during clonal expansion (Weng *et al.*, 1997) and stem cells from a number of organs (Harley *et al.*, 1994; Newbold, 1997; Wright *et al.*, 1996b), although the degree to which telomerase maintains telomeres in these cells unknown. Adult cells expressing telomerase appear to be a subset that have the potential to divide many times. Generally, telomerase appears to be strictly repressed in normal human somatic tissue (Kim *et al.*, 1994). Cultured cells from normal somatic tissues were found to be telomerase negative. However, stable telomerase activity was detected in 98 % of cultured immortal cell populations at all stages of the cell cycle (Kim *et al.*, 1994), but repressed in quiescent cells (Holt *et al.*, 1996b)

The RNA component of telomerase

Single cell eukaryotes. The essential RNA component of telomerase was first characterised in the ciliates (Greider and Blackburn, 1985; Shippenlentz and Blackburn, 1990). The *Tetrahymena* RNA component contains the sequence 5′-CAACCCCAA-3′ which serves as a template for synthesising T_2G_4 repeats (Greider and Blackburn, 1989). Elongation by *Tetrahymena* telomerase is processive, and each T_2G_4 repeat is completed before further repeats are added (Greider and Blackburn, 1989). Cleavage within the RNA template region specifically inactivates telomerase activity, and one mutation in the RNA template caused synthesis of longer telomere sequences compared to controls *in vivo,* with repeats corresponding to the mutated sequence (Yu *et al.*, 1990). Other RNA template mutations resulted in telomerase being less precise, occasionally synthesising irregular repeats not predicted by the altered template *in vivo*. These repeats probably arose due to slippage of the growing 3′ end of a stretch of five C's in the mutated template, and were stably maintained in subsequent cells. Mutant sequences in the internal portion of the telomeres caused no adverse effects as

long as wild-type sequence was present at the terminus (Yu and Blackburn, 1991). Altering the conformation of telomerase RNA outside the template resulted in a change in the enzymatic activity (Lee and Blackburn, 1993).

Yeast. The *S. cerevisiae* telomerase complex contains at least two functionally interacting RNA molecules that both act as templates for DNA polymerisation, that is, at least two active sites. It exists in two distinct states: an active ribonucleoprotein complex in an elongation-competent state, and a stalled non-elongatable complex, stably bound to the telomeric DNA following polymerisation (Prescott and Blackburn, 1997). Yeast telomerase recognises the G-rich strand with high affinity and specificity (Lue and Xia, 1998). The *S. cerevisiae* telomerase RNA component (TLC1) contains the motif 5′-CACCACACCCACACAC-3′, predicted to be the template for adding yeast telomere repeats (Singer and Gottschling, 1994). Certain mutations in TLC1 resulted in an active mutant telomerase which synthesised a pattern of mixed wild-type and mutant telomeric repeats, consistent with non-processive action. It is possible that *S. cerevisiae* telomerase stalls frequently (Cohn and Blackburn, 1995; Singer and Gottschling, 1994), and it appears to act non-processively *in vivo*, predominantly polymerising only a single (often incomplete) round of telomere elongation, rarely utilising the entire template domain (Prescott and Blackburn, 1997). Other template region mutations allowed maintenance of shortened telomeres *in vivo,* but altered specific enzymatic properties of telomerase *in vitro.*

Humans. The RNA subunit of the human telomerase complex demonstrated specific binding to a T_2AG_3 probe *in vitro* in an RNA template-dependent manner (Feng *et al.*, 1995). Two distinct complexes were detected suggesting that there is more than one form in mammalian cells (Ramakrishnan *et al.*, 1997). Cloning of the human telomerase RNA component (hTR) showed that the template region encompasses 11 nucleotides, 5′-CUAACCCUAAC-3′, complementary to the telomere sequence (T_2AG_3). Removal of hTR destroyed telomerase activity completely, and cell lines with mutations in the hTR template generated predicted mutant telomerase activity. No identity was found between the primary RNA sequence of hTR and that of the ciliates or *S. cerevisiae* (Lingner *et al.*, 1994; Romero and Blackburn, 1991). Telomerase activity does not parallel telomerase RNA (hTR) levels in human tissues: small amounts of hTR were found to be present in primary cells, normal adult somatic cells and tissues without detectable telomerase activity. Germline tissues with telomerase activity expressed high levels of hTR as predicted (Feng *et al.*, 1995).

Other proteins associated with telomerase

Single cell eukaryotes. In *Tetrahymena*, two proteins of 80 and 95 kDa co-purify with telomerase activity and telomerase RNA. The proteins are not highly related to each other or to other known sequences (Collins *et al.*, 1995). p80 is closely associated with the telomerase RNA subunit and may have catalytic activity; p95 is involved in DNA recognition (Collins *et al.*, 1995; Harrington *et al.*, 1995). Purification of the *Euplotes aediculatus* telomerase identified two protein subunits, p123 and p43. p123 contains reverse transcriptase motifs and encodes the *Euplotes* catalytic subunit of telomerase (Lingner *et al.*, 1997)

Yeast. *S. cerevisiae* strains defective in the EST (Ever Shorter Telomeres) genes EST1, EST2, EST3 and EST4 all exhibit a phenotype of progressively shorter telomeres and senescence (Lendvay *et al.*, 1996; Lundblad and Szostak, 1989), the same phenotype identified in strains deleted for the telomerase RNA component (TLC1). This suggests that the four EST genes function in the same telomere replication pathway as TLC1, and may encode either components of telomerase or factors that positively regulate telomerase activity (Lendvay *et al.*, 1996; Lin and Zakian, 1995). TLC1 is of course essential for telomerase activity (Cohn and Blackburn, 1995), however EST1 is not essential and est1$^-$ mutant strains have detectable telomerase activity (Virta-Pearlman *et al.*, 1996). The Est1p protein binds specifically to yeast G-rich telomere single stranded DNA and requires a free 3′ terminus. Est1p may mediate recognition of the chromosome end by telomerase, and is physically associated with TLC1. Changes in telomere sequence affect Est1p binding considerably (Steiner *et al.*, 1996). The EST2 gene shows homology to a conserved reverse transcriptase motif and the *Euplotes* p123 gene and encodes the integral catalytic subunit of telomerase in *S. cerevisiae*. It is required for enzyme catalysis, and mutations in just one residue result in the absence of telomerase activity and telomere shortening (Counter *et al.*, 1997; Lendvay *et al.*, 1996). EST3 encodes a novel protein (Morris and Lundblad, 1997). EST4 is a novel allele of the essential *S. cerevisiae* protein cdc13p. Cells lacking cdc13p arrest in the G2 phase of the cell cycle by a RAD9-dependent process (Weinert and Hartwell, 1993). Cdc13p binds specifically to the single stranded (TG_{1-3}) tails *in vivo* (Lin and Zakian, 1996), and cells without cdc13p contain long stretches of single-stranded DNA in telomeric and subtelomeric regions (Garvik *et al.*, 1995) and a senescent phenotype (Nugent *et al.*, 1996). It is likely that the EST4 gene is a component of telomeric chromatin (Lendvay *et al.*, 1996) and may form a docking site for telomerase at the end of the chromosome such that loss of this activity results in a telomerase-negative phenotype (reviewed Lundblad and Wright, 1996).

Humans. The human gene hTRT (also called hEST2, hTCS1 hTERT; Meyerson *et al.*, 1997; Nakamura *et al.*, 1997; Kilian *et al.*, 1997; Harrington *et al.*, 1997) encodes the catalytic subunit of human telomerase and shares significant sequence similarity to the telomerase catalytic subunit genes of lower eukaryotes (p80, p123) and *S. cerevisiae* (EST2) (Meyerson *et al.*, 1997). It is expressed at high levels in tissues expressing telomerase, and some normal tissues including thymus, testis and intestine. However, the transcript is undetectable in most other normal somatic tissues and normal cell lines which are all telomerase negative (Meyerson *et al.*, 1997). The mammalian TP1 (Telomerase-associated Protein 1) gene shows extensive amino acid similarity to the *Tetrahymena* protein p80, and interacts specifically with mammalian telomerase RNA *in vivo*. Human and murine tissues and cell lines showed widespread expression of TP1 mRNA, and often contained two transcripts (Harrington *et al.*, 1997). In the rat, the homologous TLP1 gene produces the p240 protein, which is modified to p230 *in vivo*, the dominant form in telomerase positive cells. Modification of the TLP1 protein may regulate telomerase activity *in vivo* (Nakayama *et al.*, 1997). These proteins may be components of, or closely associated with, telomerase in their respective species.

Telomerase-independent telomere maintenance

A subset of *S. cerevisiae* cells were identified that were able to maintain their telomeres without telomerase. Most est1$^-$ mutant cells in culture died, but a minor sub-population occasionally survived as a result of the acquisition and amplification of subtelomeric Y′ elements with ($TG_{1\text{-}3}$) tracts. This process does not involve telomerase activity but occurs via an alternative pathway of telomere maintenance (Lundblad and Blackburn, 1993). Non-reciprocal recombination at the boundary between telomeric and non-telomeric DNA can also result in acquisition of telomere repeats by very short telomeres (Wang and Zakian, 1990b). A similar mechanism has been identified in *Kluveromyces* yeast in telomerase RNA mutants, called telomere Cap-Prevented Recombination (CPR) (McEachern and Blackburn, 1996). Some human cell lines without telomerase activity are able to maintain telomeres at varying lengths by a telomerase-independent mechanism. A transformed human cell line without detectable telomerase activity displayed both gradual and rapid increases in telomere length that sometimes involved many kilobases. The telomere lengthening is thought to occur via a non-reciprocal recombination process with other telomeres (Murnane *et al.*, 1994; McEachern and Blackburn, 1996). Similar events have been detected in cultured cells from a Li-Fraumeni syndrome patient, heterozygous for a p53 mutation. A small number of Li-Fraumeni cells immortalised spontaneously *in vitro* - a very rare event in normal human cell populations. At crisis, the fibroblasts showed a

significant increase in the amount of telomeric DNA present, despite the lack of telomerase expression (Rogan *et al.*, 1995). Other immortalised cell lines without detectable telomerase activity were shown to have very long heterogeneous telomeres of up to 50 kb. Stabilised telomeres were achieved by an unidentified mechanism - a telomerase independent pathway termed Alternative Lengthening of Telomeres (ALT; Bryan *et al.*, 1995). This process was also seen in *S. cerevisiae* cells and termed Telomere Rapid Deletion (TRD; Li and Lustig, 1996).

Maintenance of telomeres without telomerase may not be so surprising given that some insect and plant species do not use telomerase as their primary mechanism for maintaining telomere length, such as *Drosophila,* maintaining their genome stability via transposable elements (reviewed Mason and Biessmann, 1995). These alternative mechanisms may be a back-up control to confer genome stability in the absence of telomerase.

Telomeres, cellular senescence and disease

Cellular senescence

Aging is a multifactoral process which includes macromolecular damage (Martin *et al.*, 1996), changes in protein activity (Dice, 1987) and is affected by lifestyle, disease and environmental factors. There are many genetic changes associated with aging in mammals such as de-repression of inactive genes (Wareham *et al.*, 1987), increase in gross chromosomal aberrations and spontaneous somatic mutation frequency (Dice, 1987; Bender *et al.*, 1989). Cellular senescence *in vitro* is characterised by a number of morphological changes and a slowing of the cell cycle (Campisi, 1996).

Aging in yeast occurs as a function of the number of divisions a mother cell has undergone (Kennedy *et al.*, 1994). Primary human cells in culture also have a finite lifespan (Hayflick, 1965), and undergo a limited number of population doublings before reaching senescence and ceasing to divide (Hayflick and Moorehead, 1961), although remaining metabolically active (Campisi, 1996). This phase is called the M1 growth arrest stage or checkpoint (Allsopp *et al.*, 1992). Normally at the M1 checkpoint cells permanently exit the cell cycle. In normal human somatic cells, gradual shortening of telomeres is thought to activate the complex cascade of molecular events of cellular senescence and play a role in signaling the M1 checkpoint arrest (reviewed Harley, 1991).

The replicative capacity of fibroblasts *in vitro* was found to be proportional to TRF length, and there is evidence for a critical telomere length in humans of 5-7.6 kb

(Allsopp and Harley, 1995). The lifespan of hybrid cells with elongated telomeres is longer than control cells, providing direct evidence for the involvement of telomere length in determining proliferative capacity of human cells (Wright *et al.*, 1996a). Telomere shortening is accelerated in patients with premature aging syndromes, such as Hutchinson-Gilford progeria, and cells have reduced division potential compared to age-matched controls (Allsopp *et al.*, 1992). Downs syndrome patients lose telomeric DNA faster in lymphocytes compared with age-matched controls and this is probably related to early senescence of the immune system and other cells, and the incidence of leukemia's in these patients (Vaziri *et al.*, 1993). In senescing cultures chromosomal aberrations are common, in particular end-to-end fusions and dicentric chromosome formation (Counter *et al.*, 1992), and this may be a consequence of shortened telomeres. Ataxia Telangiectasia (AT) is an autosomal recessive disorder with an increased predisposition to some cancers and premature aging (Shiloh, 1995), and lymphocytes from AT patients show accelerated telomere loss (Metcalfe *et al.*, 1996). AT cell lines show deficiency in cell cycle arrest following DNA damage (Lu and Lane, 1993) and increased levels of chromosome breakage, rearrangement and loss despite robust telomerase activity (Shiloh, 1995), and telomere length comparable to normal cell lines (Pandita *et al.*, 1995). The causative gene is homologous to the *S. cerevisiae* gene TEL1 (Greenwell *et al.*, 1995). The AT phenotype is very similar to the yeast TEL1 mutant phenotype (Greenwell *et al.*, 1995; Lustig and Petes, 1986), and may be due to a defect in telomere structure. Study of these genes and their associated phenotypes may give further clues to the structure of telomeres and their role in senescence.

There are a number of possible mechanisms for telomere length signaling cell cycle arrest. The extreme shortening of one telomere may induce a DNA damage signal such that the shortened telomere is recognised as a broken end, and further division is prevented (Wright and Shay, 1995). In *S. cerevisiae* loss of only one telomeric end leads to cell cycle arrest (Sandell and Zakian, 1993). Telomere shortening may cause dissociation of telomeres from the nuclear matrix, also signaling cell cycle arrest (Allsopp, 1996). Alternatively, genes regulating M1 could be located in the subtelomeric DNA, and changes in the heterochromatin of a short telomere may initiate senescence by altering expression of telomere-associated signaling genes (Wright and Shay, 1995). This seems unlikely given the apparent lack of PEV in human cells previously discussed (Sprung *et al.*, 1996). Other factors are strongly involved in the M1 growth arrest such as the p53 tumour suppressor gene, a cell-cycle checkpoint that can be induced by DNA damage including DNA breaks, and the Retinoblastoma tumour suppressor gene (Nelson and Kastan, 1994; Shay *et al.*, 1991a; Wright *et al.*, 1989).

Transfection of small DNA tumour viruses or their transforming genes into human cells *in vitro* extends their growth potential beyond that of normal cells by overcoming the M1 checkpoint control (Counter *et al.*, 1992; Shay *et al.*, 1991b; Shay and Wright, 1989; Shay *et al.*, 1993; Van der Haegen and Shay, 1993; Wright *et al.*, 1989). However, during this period there is continued telomere shortening (Counter *et al.*, 1992; Shay *et al.*, 1993), until the average telomere length is approximately 1.5 kb (Counter *et al.*, 1992; Shay *et al.*, 1993). After additional doublings, M1 cells enter crisis, a period controlled by an independent M2 stage (Shay *et al.*, 1991a) where cell death occurs (Shay and Wright, 1989; Stein, 1985; Wright *et al.*, 1989), probably due to genome instability.

In order for cells to overcome senescence and become immortal they must escape the M1 and M2 checkpoints that limit the proliferative capacity of normal cells. Hybrid studies have shown that overcoming senescence requires recessive mutations in growth regulatory genes (Pereira-Smith and Smith, 1988). Once immortal, cells can follow many pathways that result in the progression of cancer (Nowell, 1986). Progression beyond M2 crisis is a very rare event. Selection for mutations which allow some chromosomal stability will allow immortal clones to survive with stabilised genomes, and unlimited proliferative capacity to form tumours (Counter *et al.*, 1992; Counter *et al.*, 1994a). In immortal cell cultures, telomere length and the frequency of dicentric chromosome formation stabilised after M2 crisis (Counter *et al.*, 1992). It was proposed that telomerase reactivation is an essential requirement for cellular immortality to confer genome stability (Wright and Shay, 1995). The 'telomere hypothesis' of cellular aging is shown in figure 1.4. An immortal cell line is defined as one that is able to double the maximum expected lifespan (Shay *et al.*, 1991a). In some human cells, telomerase expression appears to be sufficient for cellular immortalisation (Bodnar *et al.*, 1998; Vaziri and Benchimol, 1998). Introduction of the human telomerase catalytic subunit (hTRT) into telomerase-negative normal human cells resulted in expression of telomerase and cessation of telomere shortening (Bodnar *et al.*, 1998). These cells were maintained in continuous culture for 280 population doublings (controls senesced at 75-80 PD), and were still dividing without showing signs of senescence. The cells had normal karyotype stability, normal *in vitro* growth requirements, and cell cycle checkpoint controls, including p53 and pRB, were intact (Morales *et al.*, 1999). Telomerase expression appears to be sufficient for immortalisation in these cells. Ectopic expression of hTRT in post-M1 telomerase-negative cells also allows them to proliferate beyond M2 crisis with arrested telomere shortening (Counter *et al.*, 1998; Halvorsen *et al.*, 1999). Ectopic expression of telomerase in cell culture or *in vivo* may be useful in extending the life-span of cells, with potential applications in biotechnology (Bodnar *et al.*,

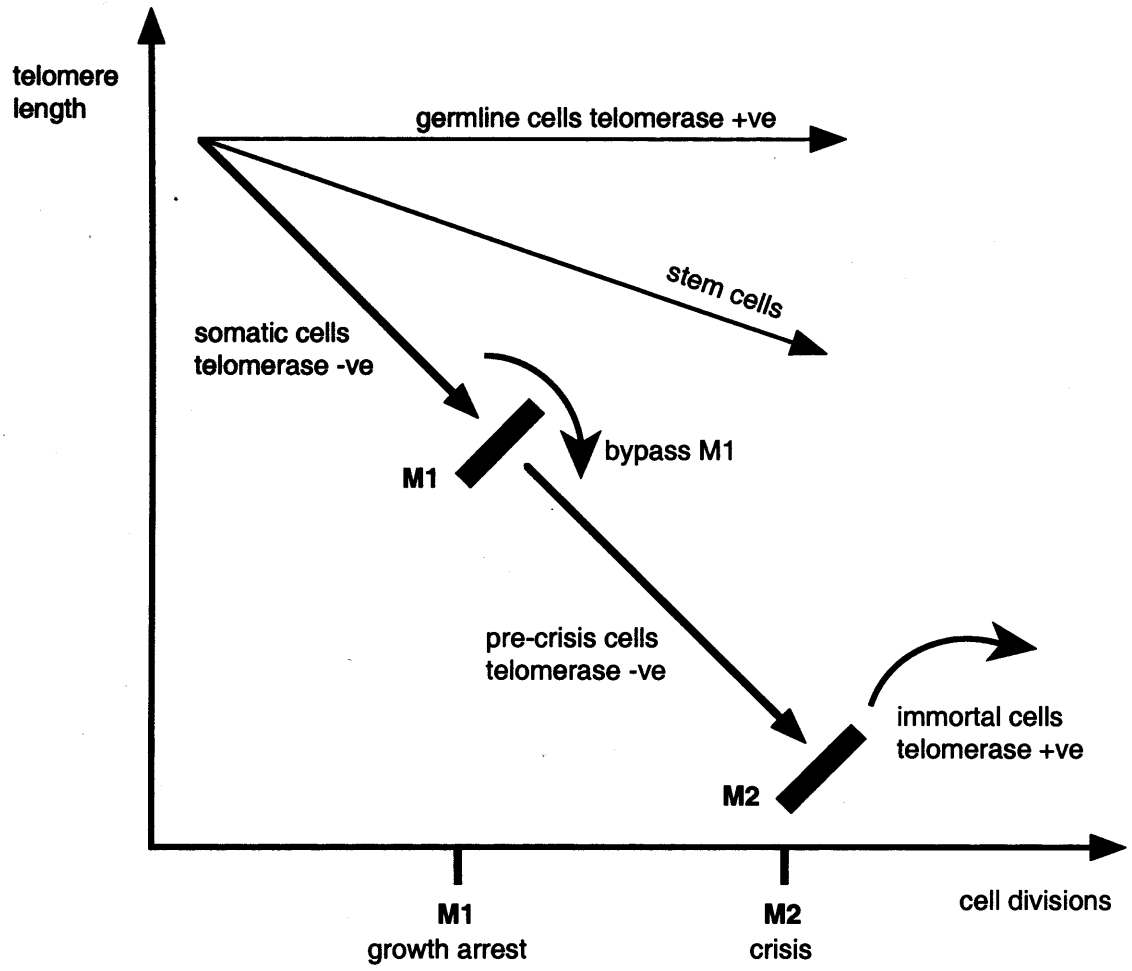


Figure 1.4 The telomere hypothesis of cellular ageing and immortalisation

Telomeres shorten with every cell division in normal somatic cells until they reach the M1 growth arrest stage and enter senesence. Critical telomere shortening is thought to signal cell cycle arrest at M1. Stem cells express low levels of telomerase and have extended proliferative capacity. Their telomeres still shorten, although at a slower rate than somatic cells. Transfection with tumour viruses and consequent loss of tumour suppressor genes or activation of oncogenes allows cells to bypass M1 with an extended lifespan, and telomere shortening continues. Cells enter M2 crisis when the genome becomes critically unstable, and most cells die. Cells with mutations allowing genome stabilty can escape M2 and become immortal. Evidence suggests that telomerase reactivation is required for cellular immortalisation. Figure adapted from Harley *et al.*, 1994.

1998), gene therapy and treatment of age-related disease (Vaziri and Benchimol, 1998).

Telomerase and Cancer

Cancer mainly arises from the accumulation of several independent mutations in a cell, conferring a growth advantage (Nowell, 1986). This requires repeated expansion of mutant clones which will almost certainly need to become immortal at some point in order to allow enough divisions to form tumours. It is thought that the process of cellular senescence may be an anti-cancer mechanism, preventing cells from undergoing enough clonal expansions to form a tumour by having their growth halted after a certain number of divisions (Wright and Shay, 1995). It was hypothesised that telomerase reactivation is a requirement in tumour cells in order to escape the constraints of the M1 and M2 checkpoints (Shay and Wright, 1996). Short but stable telomeres are common in many different tumour types (Hastie *et al.*, 1990; Counter *et al.*, 1994b; Broccoli *et al.*, 1995; reviewed Bacchetti and Counter, 1995) although elongated telomeres have been seen in some colorectal carcinomas (Hastie *et al.*, 1990; reviewed Bacchetti and Counter, 1995) Assays on a variety of cancers showed telomerase activity in 90 % of tumour samples but not in adjacent normal or benign tissues (Counter *et al.*, 1994b; Kim *et al.*, 1994). Consequently, telomerase inhibition in malignant cells is a target for specific anti-cancer therapy (Harley *et al.*, 1990).

However, it is unclear at what stage telomerase is reactivated, and whether it is a requirement for progression to malignancy. Cells in culture immortalised by ectopic telomerase expression (Bodnar *et al.*, 1998) did not show any changes typically associated with malignant transformation even after inactivation of their p53 and pRB pathways (Morales *et al.*, 1999), and did not form tumours in nude mice (Jiang *et al.*, 1999). It appears that the presence of telomerase activity merely indicates that a cell is able to inactivate the telomere shortening which limits proliferative capacity of normal somatic cells, conferring potential immortality but not malignancy (Holt *et al.*, 1996a). Some cancers without telomerase activity have been identified, for example some cases of chronic lymphocyte leukaemia (Counter *et al.*, 1995) and Retinoblastoma (Gupta *et al.* unpublished, cited in Bacchetti and Counter; 1995). These tumours may be able to maintain telomere stability without telomerase activity, as seen in some immortal cell lines (Bryan *et al.*, 1995; Murnane *et al.*, 1994; Rogan *et al.*, 1995). A subset of neuroblastoma stage IV tumours (IVS) expressed telomerase activity in only 36 % of samples studied (Hiyama *et al.*, 1992; Hiyama *et al.*, 1995; Kim *et al.*, 1994) Interestingly, these tumours often regress independently of anticancer therapies (Evans *et al.*, 1971; Hiyama *et al.*, 1995) possibly reflecting genome instability and

cell death due to lack of telomere maintenance (Hiyama *et al.*, 1995). Some cells may accumulate only a few mutations before becoming malignant without reaching the normal limit of division and would therefore be mortal and telomerase negative (Holt *et al.*, 1996a). Many malignancies are thought to arise from stem cells (Potten and Loeffler, 1990; Greaves, 1996; Pierce and Speers, 1988), and it is possible that telomerase expression in tumours is purely a marker of telomerase-positive stem cell origin, which by nature require an extended replicative capacity (Greaves, 1996). Descendants of stem cells can have reduced or absent telomerase activity associated with partial differentiation, and this may explain the occurrence of tumours with short telomeres or low telomerase activity (Counter *et al.*, 1995; Greaves, 1996).

A mouse model for human telomerase activity

An $mTR^{-/-}$ knockout mouse lacking detectable telomerase activity was unexpectedly viable for six generations studied. However, some $mTR^{-/-}$ embryos developed abnormally (Herrera *et al.*, 1999), and later generations exhibited defective spermatogenesis and impaired proliferative capacity of haematopoietic cells (Lee *et al.*, 1998). $mTR^{-/-}$ cells under long term culture conditions displayed continual telomere shortening, and a gradual reduction in growth rate until approximately 450 divisions, when cell growth almost stopped, suggesting that telomerase-dependent telomere maintenance is essential for cell growth (Niida *et al.*, 1998). Telomerase deficient cells from $mTR^{-/-}$ mice immortalised by viral oncogenes were able to generate telomerase-negative tumours in nude mice, indicating that telomerase activity is not an essential component for malignancy in murine cells (Blasco *et al.*, 1997).

However, in contrast to human adult tissues, most mouse adult tissues express telomerase (Kim *et al.*, 1994; Broccoli *et al.*, 1995; Prowse and Greider, 1995; Wright *et al.*, 1996b). The murine telomerase RNA component, mTR, was detected in testis, intestine, liver and spleen and correlated with telomerase expression in these tissues, contrasting with adult human cells which express hTR but do not have telomerase activity (Feng *et al.*, 1995). This discordance in expression and activity suggest that murine and human telomerases are under different levels of control or are biochemically different. This may reflect the ease of primary mouse cells to immortalise spontaneously in culture (10^6 times more frequently) relative to human cells (Ponten, 1976; Macieiracoelho and Azzarone, 1988; McCormick and Maher, 1988). These differences suggest that the mouse is a poor model for the study of human telomerase expression and activity.

Proterminal and subterminal regions of chromosomes

Isolating telomeres and adjacent regions

Telomeres have been isolated by digesting genomic DNA with restriction enzymes that do not cut in $(T_2AG_3)n$ repeat arrays, then size fractionating to enrich for tandemly repeated sequences including telomeres, and ligating these fragments into plasmid vectors (de Lange *et al.*, 1990). This method resulted in the isolation of some clones containing human subtelomeric sequence adjacent to telomere repeats. Telomeres can be isolated more successfully using Yeast Artificial Chromosome (YAC) vectors containing a yeast selectable marker, an autonomously replicating sequence, a yeast centromere and an array of *Tetrahymena* telomere repeats. The YAC is ligated to digested human DNA enriched for telomeres. Half-YACs containing human telomere repeats at one end can only survive in a yeast cell if 'rescued' by the addition of yeast telomere repeats onto the human and *Tetrahymena* telomere repeats. This method has been used to isolate small recombinant YACs containing several kilobases of DNA from human chromosome ends (Brown, 1989; Cross *et al.*, 1989) and YACs with over 200 kb of telomeric DNA (Riethman *et al.*, 1989). PCR-based techniques have also been effective in isolating sequences from the proximal ends of telomeres and short sequences adjacent to telomeres (Weber *et al.*, 1990; Royle *et al.*, 1992). Fragments larger than 5.5 kb resistant to *Mbo*I digestion (which does not cut within telomere repeat arrays) were selected and ligated to linker-amplimers. PCR with a telomere primer and a linker primer resulted in products amplified from the proximal ends of telomeres into the adjacent DNA (Royle *et al.*, 1992). This method is quicker and easier than the conventional YAC cloning strategy.

Proterminal regions

Proterminal regions are described as the distal regions of human chromosomes (reviewed Royle, 1995), and have a high density of CpG island clusters and genes (Craig and Bickmore, 1994; Saccone *et al.*, 1992). These regions support homologous chromosome pairing during meiosis (Wallace and Hulten, 1985) and are presumably involved in the initial homology searches preceding chromosome synapsis, which begins near the telomere (Rasmussen and Holm, 1980). Proterminal regions also support high levels of meiotic recombination (Donis-Keller *et al.*, 1987; NIH/CEPH, 1992; Weissenbach *et al.*, 1992), indicated by high recombination frequencies and expansion of the genetic linkage-map distances relative to the physical map, especially in male meiosis (Rasmussen and Holm, 1980; Donis-Keller *et al.*, 1987; Rouyer *et al.*, 1990). Minisatellites are tandem repeat arrays with repeat units between 9 and 90 bases (Armour *et al.*, 1990; Wong *et al.*, 1987) and lengths from 0.5 to 30 kb (Jeffreys, 1987). These repeats can be highly

variable, with heterozygosity at some loci exceeding 90 % (Wong *et al.*, 1987). GC-rich minisatellites are clustered in proterminal regions (Royle *et al.*, 1988) and are often associated with dispersed repeat elements such as *Alu*, L1 and retroviral LTR-like sequences (Armour *et al.*, 1989). Many minisatellites show extreme variability in allele length and internal structure of the repeat array, and mapping internal structures of hypervariable minisatellite repeat arrays has identified a high germline mutation rate and complex gene conversion-like events of mutation (Jeffreys *et al.*, 1994; May *et al.*, 1996). Hypervariable minisatellites may be 'hotspots' for recombination or conversion events, initiated in the DNA flanking the minisatellite (Jeffreys *et al.*, 1994).

Subterminal regions

Yeast. Subterminal regions are the distal ends of chromosomes immediately adjacent to telomeres. Yeast chromosome ends are dynamic in terms of subtelomeric repeat structure and variability. They are composed of several different repeat elements, the main ones being X and Y′. The pattern of X and Y′ element distribution on individual chromosomes is different between strains of *S. cerevisiae* (Zakian and Blanton, 1988). The X element comprises at least four sequences (45-140 bp) and a possible 'core' element of 560 bp found at some chromosome ends. The Y′ element is found immediately adjacent to the (TG_{1-3}) telomere sequences on some chromosomes. Y′ elements vary in copy number and location between strains, and a single Y′ is able to expand into a tandem array by unequal sister chromatid exchange (Horowitz *et al.*, 1984; Louis and Haber, 1990). There are sometimes stretches of (TG_{1-3}) between X and Y′ elements, creating proximal and distal domains (Chan and Tye, 1983; Louis *et al.*, 1994). Subterminal sequences can be absent from chromosome ends and are therefore probably not essential components for chromosome stability (reviewed Zakian, 1995).

Humans. Human subterminal regions are usually composed of several members of low copy repetitive sequence families organised as a string of distinct repeated sequence elements. Subtelomeric regions often contain short interstitial telomere-like arrays of T_2AG_3 repeats (Wells *et al.*, 1990). In humans, distinct subterminal sequence families are located mainly, but not always, at telomeres on most or all autosomes. Chromosome distribution and copy number of subterminal repeat families are polymorphic between unrelated individuals. This polymorphism probably arises by exchange between different telomeres, which has been shown to occur in yeast (Louis and Haber, 1990; Louis and Haber, 1992; Murray and Szostak, 1983). Some chromosomes do not cross-hybridise to these sequence families, such as Xp and Yp (Brown *et al.*, 1990; Royle *et al.*, 1992) and 7q (Brown *et al.*, 1990; Riethman *et al.*, 1989), and some subterminal sequence families

are unique in the genome. Examples of these subterminal sequences include TelBam3.4, TelSau2.0 (Brown *et al.*, 1990), pTH2, pTH14 (de Lange *et al.*, 1990), Tsk6 and Tsk37.1 (Royle *et al.*, 1992), all of which show homology to one another and constitute a family of subterminal sequences. These repeats are often truncated variably before the start of the telomere, and this may indicate that telomere loss and healing can occur in this repeat family (Royle *et al.*, 1992). Other examples include TelBam11 (Brown *et al.*, 1990), pHutel-2-end (Cross *et al.*, 1990), Tsk7 and Tsk48.1 (Royle *et al.*, 1992). More than 95 % of these subtelomeric sequences are localised to within 30 kb of chromosome ends (Cross *et al.*, 1990).

A large number (661) of sequence tagged sites (STS) from the terminal 300 kb of 31 human chromosome ends have been isolated, and approximately 30 % of the sequence (excluding *Alu*, L1 and other repetitive sequences) was present on more than one chromosome end (Rosenberg *et al.*, 1997). The presence of similar subtelomeric sequences at more than one telomere may lead to telomere-telomere exchange by recombination between non-homologous chromosomes. This may explain why subsets of telomeres detected by some repeats families are polymorphic within a population (Brown *et al.*, 1990; Cross *et al.*, 1990; IJdo *et al.*, 1992). Comparison of DNA sequence from the ends of three human chromosomes, 4p, 16p and 22q, showed that the subterminal regions can be subdivided by degenerate T_2AG_3 repeats into distal and proximal sub-domains, with different patterns of homology to other chromosome ends. The distal regions contain a patchwork of numerous short (<2 kb) segments of interrupted homology to many other human ends. The proximal regions show much longer (10-40 kb) uninterrupted homology to only a few chromosome ends (Flint *et al.*, 1997). The lack of highly conserved structure or sequence in human subtelomeric regions suggests that they act only as a buffer against the effects of breakage rather than being an essential chromosomal component, and chromosomes lacking the 16p subtelomeric region have been shown to replicate and segregate normally (Flint *et al.*, 1994). Subtelomeric sequence is also found in other species, but is species-specific. Organisation of sequences adjacent to telomeres is very different even in closely related primates like chimpanzee, orangutan, gorilla and human. For example, a subterminal satellite (32 basepair AT rich repeat) found in the chimpanzee genome hybridises to twenty-one AT-rich terminal bands and two interstitial sites (7q22.2 and 13q14.2). The satellite is more abundant in gorillas but not detected in humans or orangutans (Royle *et al.*, 1994).

Length polymorphism at the 16p telomere

A long range restriction map of terminal 16p13.3 revealed major polymorphic length variation in this region. This region is well characterised due to the

presence of the α-globin gene cluster. Four alleles, A, B, C and D, were found, where the α-globin locus lies 170 kb, 350 kb, 430 kb or 245 kb from the telomere, indicating that 16p varies in length by up to 260 kb. Allele A represents 70 % of all 16p chromosomes (Harris and Higgs, 1993; Harris and Thomas, 1992; Wilkie *et al.*, 1991). The two most common alleles, A and B (24 %), contain different terminal segments containing sequences related to different chromosome termini, starting 145 kb distal to the α-globin genes. In a 4.5 kb segment proximal to the divergence point, 16p alleles A and B show reduced homology. Uncommon alleles C and D are related, but not identical, to the B allele beyond the divergence point and probably result from more distal exchange events or truncations involving the B allele (Harris and Higgs, 1993). The 16p polymorphism is found in rare racial groups, and therefore probably occurred before human radiation, with more distal exchanges occurring more recently (Wilkie *et al.*, 1991). The end of the long arm of chromosome 16 (16q) also demonstrates length polymorphism, with three alleles varying in length by up to 125 kb (Harris and Thomas, 1992). A possible 55 kb chromosome length polymorphism at the 2q telomere has also been identified (Macina *et al.*, 1994). Given this evidence, length polymorphism is likely to be present at other chromosome ends, and may indicate exchanges of terminal DNA between non-homologous chromosomes (Wilkie *et al.*, 1991).

Chromosome breakage and healing

Developmentally programmed chromosome healing

During the formation of a new macronucleus in the ciliate *Euplotes crassus*, micronuclear chromosomes are reproducibly broken at around 10,000 sites. This is coupled with *de novo* telomere synthesis by telomerase to generate short linear macronuclear DNA molecules (Yu and Blackburn, 1991; reviewed Zakian *et al.*, 1990). Chromosome fragmentation/telomere addition sites (E-Cbs) are conserved sequence elements, found at one or both ends of most macronuclear DNA molecules (Yao *et al.*, 1987). In *Tetrahymena thermophila*, the Cbs element is required for chromosome fragmentation but shows little sequence similarity to E-Cbs (Klobutcher *et al.*, 1998). In presomatic cells of the early embryo in nematode *Ascaris lumbricoides*, a process termed chromosome diminution takes place, involving elimination of 25 % of germline-specific DNA and heterochromatic chromosome ends. Fragmentation of the rest of the chromosome occurs to make multiple new chromosomes. Breakage takes place in a short specific region called CBR at many different sites. The newly formed ends of the reduced somatic chromosomes carry tandem repeats of the telomeric sequence T_2AG_2C. These telomeric repeats are not present at these sites in the germline genome and appear to be added *de novo* by telomerase (Muller *et al.*, 1991).

Spontaneous chromosome breakage and healing

Programmed breakage and healing events do not take place in the mammalian genome. However, the introduction of cloned telomeric DNA into mammalian cell lines can lead to integration into chromosomes and seed formation of new telomeres, probably by telomerase (Farr *et al.*, 1991). In human cells, a few large truncations were formed, however, most of the integration events appeared to be close to the pre-existing ends of natural chromosomes generating small truncations too small to detect by cytogenetic methods (Barnett *et al*,. 1993). Introduction of non-telomeric plasmid DNA into human cell lines occasionally resulted in chromosome breakage, integration and (T_2AG_3) repeat addition to non-telomeric DNA, probably by telomerase, although this was a rare event (Murnane and Yu, 1993). Generally mammalian cell lines appear to restrict telomere sequence addition to telomere-like repeat arrays or close to existing ends of chromosomes, presumably because promiscuous healing of breaks in the germline cells would lead to karyotype instability (Farr *et al.*, 1991). It is possible that cells with small cryptic terminal truncations are more likely to survive than a cell which has suffered a massive deletion (reviewed Farr, 1996).

Telomerase is thought to heal and stabilise chromosomes broken in the human germline, as well as maintaining telomere length. This mechanism potentially gives rise to chromosomes carrying terminal deletions and a new telomere. It is not known how frequent breakage and healing events are resulting in terminal deletions. There are examples of chromosome healing by direct telomere addition, possibly by telomerase, *in vivo*. Rearrangements involving small interstitial deletions at the tip of the short arm of chromosome 16 (16p13.3) usually involve the α-globin gene cluster, and are therefore easily identified by the phenotype of α-thalassaemia (reviewed Higgs *et al.*, 1989). Other groups of rearrangements at 16p consist of larger deletions of 1-2 Mb, resulting in α-thalassaemia and mental retardation syndrome (ATR-16) (Wilkie *et al.*, 1990). The clinical features of ATR-16 appear to be associated with monosomy of 16p13.3. One patient with ATR-16 has a *de novo* truncation of the terminal 2 Mb of 16p which appears to have been stabilised *in vivo* by direct addition of telomere repeats (Lamb *et al.*, 1993). Similar truncations were found in six other individuals with α-thalassaemia or ATR-16 resulting from deletions of 16p13.3, and they also appear to have been stabilised by direct addition of T_2AG_3 repeats to non-telomeric DNA (Flint *et al.*, 1994; Lamb *et al.*, 1993). In all but one of these cases, the 3-4 nucleotides at the breakpoint were complementary to and in phase with the RNA template of human telomerase. It was not possible to determine exactly at which nucleotide the *de novo* addition of repeats occurred as the 3-4 nucleotides at the breakpoints also corresponded to the normal chromosome. At the remaining breakpoint, two extra nucleotides had been

inserted between the end of the normal sequence and the *de novo* repeat array. It is highly likely that telomere addition at these breakpoints was mediated by telomerase. A similar truncation has been found on chromosome 22. The patient has overlapping features of the cytogenetically visible 22q.13.3 deletion syndrome, including mild mental retardation and delay of expressive speech. The breakpoint also appears to have been healed by the direct addition of a telomere, possibly by telomerase (Wong *et al.*, 1997).

It is not clear how telomerase might recognise broken ends or begin to add telomere repeats. Using an *in vitro* assay, human telomerase was shown to correctly recognise an ATR-16 breakpoint sequence and add T_2AG_3 repeats, generating the correct junction sequence (Morin, 1991). This suggests that 3′ termini with only very minimal complementarity to the RNA template of telomerase are sufficient for recognition and repeat addition (Morin, 1991). No other common structural features were identified at the breakpoints that may have facilitated recognition by telomerase. If this type of healing event is a frequent occurrence, it is surprising that similar events have not been seen more often. The apparent frequency of these events at the tip of 16p may be due to the phenotype, which provides a distinctive marker for rearrangements, including deletions in this region, and the relatively terminal location of the α-globin genes, resulting in only small deletions. Terminal deletions of other chromosomes may be either less distinct, and are therefore not detected, or more deleterious, resulting in loss from the cell population (reviewed Broccoli and Cooke, 1993).
However, addition of a new telomere is not the only mechanism by which chromosomes can be healed. A subtelomeric rearrangement involving 16p has been identified, giving rise to α-thalassaemia by deletion of a major remote regulatory element controlling gene expression. The breakpoint lies within an *Alu* element about 105 kb from the 16p subtelomeric region. The broken chromosome has been stabilised by 'telomere capture' due to recombination with a closely related *Alu* element on chromosome 16. This rearranged chromosome appears to have been transmitted normally within the family studied and may reflect a more general process by which subtelomeric sequences are normally dispersed between chromosome ends (Flint *et al.*, 1996).

Terminal deletion syndromes

Deletions can range from one base pair to several megabases of DNA. Common mechanisms of deletion include chromosomal breakage, parental translocation or inversion (due to chromosome breaks) and unequal crossing over (reviewed Connor and Ferguson-Smith). Unequal crossing over is more likely to occur in regions with duplicated genes or repeated regions. Deletion of an autosomal

region commonly leads to mental retardation, multiple congenital malformations, dysmorphic features and growth retardation, probably due to haplo-insufficiency of some genes. Certain deletions give typical clinical phenotypes enabling diagnosis, but some clinical features may be non-specific, making identification of the deleted chromosome difficult if it is not cytogenetically visible. Deletion syndromes (contiguous gene disorders) often result from deletions removing a number of genes causing more than a single gene disorder, and are usually sporadic, but can be familial. Microdeletions usually involve less than 5 kb and often arise via cryptic reciprocal translocations between chromosome ends, which cannot be detected using conventional cytogenetic banding techniques. Individuals with balanced translocations are unaffected, but may have unbalanced affected progeny (reviewed Royle, 1995). Terminal microdeletions have been identified on a number of chromosomes, for example small deletions at the tip of the short arm of chromosome 16 (16p) involving the α-globin gene cluster.

Miller-Dieker syndrome (MDS) has a phenotype including classical (type I) lissencephaly (smooth brain) with mental retardation, and a characteristic facial appearance. The phenotype results from microdeletions in band 17p13 with a critical region in band 17p13.3, although some patients have larger cytogenetically visible deletions. A disease gene LS1 (lissencephaly-1) has been identified (Reiner *et al.*, 1993) and it is likely that this gene is involved in a signal transduction pathway needed for cerebral development. Haplo-insufficiency of LS1 appears to cause Miller-Dieker syndrome (Chong *et al.*, 1996). MDS is thought to be a contiguous gene syndrome and it appears that deletion of additional genes distal to LIS1 is responsible for the facial dysmorphia and other abnormalities seen in MDS patients. A number of patients with MDS have been shown to have cryptic reciprocal translocations (Kuwano *et al.*, 1991). A large but cytogenetically invisible deletion due to a cryptic translocation between 17pter and 8qter was revealed by FISH (fluorescence *in situ* hybridisation) analysis using probes for the distal region of 17p. The father was a balanced carrier of t(8q:17p), and the proband inherited an unbalanced chromosome leading to hemizygosity for the tip of 17p. Another 'half-cryptic' translocation was identified between 17pter and 3pter in a patient with MDS. In this case, additional material was detected on only one chromosome (Kuwano *et al.*, 1991; reviewed Ledbetter, 1992).

Wolf-Hirschorn syndrome is caused by partial deletion of 4p and has a phenotype including severe growth retardation, mental retardation, microcephaly, a typical 'greek helmet' facial appearance and closure defects such as cleft lip. The critical region, 4p16.3, was identified by studying an inherited sub-microscopic

translocation between chromosomes 4p and 19p. The patient had a normal karyotype. However, the translocation was detected by *in situ* hybridisation with probes for the telomeric region of 4p, indicating hemizygosity for the distal region of 4p (Altherr *et al.*, 1991). The critical region is approximately 165 kb long and gene dense (Wright *et al.*, 1998). Some cases of Wolf-Hirshorn syndrome are due to inheritance of the unbalanced products of reciprocal translocations (McKeown *et al.*, 1987), other cases are due *de novo* deletion of 4p (Quarrell *et al.*, 1991; Thies *et al.*, 1992).

Detection of cryptic chromosomal rearrangements

Until recently, the detection of cryptic chromosomal rearrangements and microdeletions less than 1-2 Mb (not detected by conventional cytogenetic methods) relied on the presence of specific clinical features that would pin-point a specific chromosomal region to look at, such as α-thalassaemia (16p), Wolf-Hirschorn syndrome (4p) and Miller-Dieker syndrome (17p). Often, patients with mental retardation and dysmorphic features do not fit a specific clinical description. Rearrangements without a specific phenotype are difficult to diagnose, and consequently may be an important cause of human genetic disease (Lamb *et al.*, 1989). Mental retardation is present in about 3 % of individuals but is unexplained in more than half of these cases due to the aetiological heterogeneity of this condition. As many as 6 % of idiopathic mental retardation referrals are estimated to possess submicroscopic rearrangements involving telomeres and the associated subtelomeric DNA, including microdeletions (Flint *et al.*, 1994; Giraudeau *et al.*, 1997).

A number of different methods have been used to detect microdeletions and cryptic rearrangements. A number of terminal deletions at 16p causing α-thalassaemia were identified using DNA probes for the terminal region of 16p, and gene dosage analysis (Wilkie *et al.*, 1990). Cloning of individual human telomeres at 4p (Bates *et al.*, 1990), 7q (Riethman *et al.*, 1989; Brown *et al.*, 1990), 16p (Wilkie *et al.*, 1991) and Xp/Yp (Brown *et al.*, 1990) provided the ends for the genetic and physical maps and have been useful for generating diagnostic probes for cytogenetics (reviewed Ledbetter, 1992). Chromosome specific markers for 27 telomeres, including 149 STSs and 24 polymorphisms for proterminal regions were generated in 1996, although not all chromosome ends were represented in this resource (Vocero-Akbani *et al.*, 1996). Alternatively, FISH can be used routinely for detection of microdeletions and cryptic rearrangements, and is more sensitive than cytogenetic methods previously used (reviewed Ledbetter 1992). A strategy to screen for the abnormal inheritance of subtelomeric DNA in individuals with mental retardation but normal karyotypes detected three abnormalities in 99

patients. The strategy involved polymorphic marker analysis at 28 chromosome ends. Two cryptic translocations and one terminal microdeletion were identified and confirmed by FISH, reverse chromosome painting and pulse-field gel electrophoresis (PFGE) (Flint *et al.*, 1995). A complete set of human telomeric FISH probes is now available, enabling detection of unique sequence at each telomere, located at a known distance 100-300 kb from the end of most chromosomes. This results in a 10-fold improvement in deletion-detection sensitivity compared with high-resolution cytogenetic methods, which have a lower detection limit at 2-3 Mb (NIH/IMM, 1996). A FISH technique has been developed comprising a resource of 41 telomere-specific probes, allowing simultaneous analysis of the subtelomeric regions of every chromosome for deletion, triplication and balanced translocation events. It requires only a single microscope slide per patient and is expected to be a very useful diagnostic tool with applications in idiopathic mental retardation, congenital abnormalities and some forms of cancer (Knight *et al.*, 1997).

Project Aims

Microdeletions are potentially an important source of human genetic disease (Flint *et al.*, 1995). A number of terminal deletions on different chromosomes appear to have been healed by the direct addition of a new telomere, and it is highly likely that repeat addition was mediated by the enzyme telomerase. However, the frequency at which these healing events occur is not yet known. An anchored PCR-based strategy was used successfully to isolate a terminal breakpoint from a patient with a visible deletion on 7q (Royle NJ, personal communication). The aims of this project were to:

- improve the efficiency of the telomere-anchored PCR-based strategy used for the isolation of terminal deletion breakpoints by modification of appropriate steps.
- develop a strategy to increase the enrichment of chromosome ends adjacent to telomere repeat array from patient genomic DNA for telomere-anchored PCR.
- isolate terminal deletion breakpoints that may have been healed by telomerase-mediated addition of a new telomere from a number of patients with clinical features including mental retardation.
- compare sequence in the vicinity of the breakpoints to determine whether telomerase has specific sequence requirements for recognising broken chromosomes and adding telomere repeats.

Chapter 2

MATERIALS AND METHODS

Materials

Chemical reagents, molecular biology reagents and equipment
Chemicals were obtained from Fisher Scientific (Loughborough, UK), Fisons (Loughborough, UK), Flowgen (Stafford, UK), FMC Bioproducts (Rockland, USA), Serva (Heidelberg, Germany) and Sigma Biochemical Company (Poole, UK). Molecular Biology reagents were obtained from Advanced Biotechnologies (Leatherland, UK), Amersham International Plc (Little Chalfont, UK), Bio-Rad (Hemel Hemstead, UK), Boehringer Mannheim (Lewes, UK), Gibco BRL (Paisley, UK), National Diagnostics (Hull, UK), NEN Life Sciences (Hounslow, UK), New England Biolabs (Hitchin, UK) and Pharmacia (Milton Keynes, UK). Specialised equipment was obtained from Bio-Rad, Cecil Instruments (Cambridge, UK), Genetic Research Instrumentation (Dunmow, UK), Hybaid (Teddington, UK), and Perkin-Elmer/Applied Biosystems (Beaconsfield, UK).

Human DNA samples
DNAs from lymphoblastoid cell lines comprising the Centre d'Etude du Polymorphism Humain (CEPH, Paris, France) panel of Caucasian DNAs were obtained from Professor H. Cann and Dr. J. Dausset. The NT (Flint *et al.*, 1995), RoBa (Giraudeau *et al.*, 1997) and BQ lymphoblastoid cell lines were obtained from Dr. J. Flint (Institute of Molecular Medicine, Oxford, UK). The CB0001 and CB00054 lymphoblastoid cell lines (Wilson *et al.*, 1995), and genomic DNAs BW and DW were obtained from Professor R. Trembath (University of Leicester, UK). Lymphoblastoid cell lines AJ, CB and NS (Nesslinger *et al.*, 1994), and genomic DNAs KJ and FJ were obtained from Dr. H. McDermid (University of Alberta, Edmonton, Canada). The FB336R and FB240 fibroblast cell lines were obtained from M. Muenke (Division of Human Genetics and Molecular Biology, University of Pennsylvania School of Medicine, Philadelphia, USA; described in Gurrieri *et al.*, 1993). These cell lines are also available from the NIGMS Human Genetic Mutant Cell Repository (Camden, New Jersey, USA). Human placental genomic DNAs P50, P53 and P55 were obtained from anonymous donors at the Leicester Royal Infirmary.

Oligonucleotides

Oligonucleotides were synthesised in-house (Dr. K. Lilley, Protein and Nucleic Acids Laboratory, University of Leicester, UK). Oligonucleotides were ethanol precipitated and re-dissolved in distilled water before use. The concentration was determined by absorbence at 260 nm using a Cecil Instruments 2040 UV spectrophotometer. Sequences of oligonucleotides are shown in table 2.1. Random hexadeoxyribonucleotides for random primed labeling were obtained from Pharmacia.

TABLE 2.1 PRIMER SEQUENCES

Non-complementary tails are indicated in bold. Restriction enzyme sites are underlined.

Primer name	Primer sequence 5′ to 3′
Telomere primers	
TelB	**GGCC**<u>ATCGATGAATTC</u>GTTAGGGTTAGGGTTAGGGTTAGG
TelC	**GGCC**<u>ATCGATGAATTC</u>TAACCCTAACCCTAACCCTAA
TelG	CCCTCACCCTCACCCTCACCCTC
TelGcomp	AGGGTGAGGGTGAGGGTG
TelJ	ACCCCAACCCCAACCCCAACCCC
TelJcomp	GGGTTGGGGTTGGGGTTG
TelK	CCCTGACCCTGACCCTGACCCTG
TelKcomp	AGGGTCAGGGTCAGGGTC
Tel1	TTAGGGTTAGGGTTAGGG
Tel2	CTAACCCTAACCCTAACC
Sequencing primers	
M13forward	GTAAAACGACGGCCAGT
M13forward (21mer)	CGTTGTAAAACGACGGCCAGT
M13reverse	AACAGCTATGACCATG
M13reverse (21mer)	CACACAGGAAACAGCTATGAC
SK	TCTAGAACTAGTGGATC
T3	ATTAACCCTCACTAAAG
T7	AATACGACTCACTATAG
Linker primers	
Sau-L-A	GC<u>GGTACCCGGGAAGCTT</u>GG
Sau-L-B	<u>GATCCCAAGCTTCCCGGGTACC</u>GC

Clone specific primers	
CB000124cA	GATCAGCTTATCCACCTCTGTCC
CB000124cB	CTTATTTGTGGATTTGTAGACCAC
CB000114dA	ACAAGAGTCAGCAAGTATTTTCTGC
CB000128	GATCCTTCCTCTTTGCAGCC
R79bA	CCAAGGTCTCACTGGCACTC
R79bB	GTGATTGTTGGCGATAAGGTGC
AJ111B1	GCACTTTAAATGGGCGACAGAGG
AJ111B2	CCTAAGGGAATGAGTTAATACCAG
AJ111C	GGAATGAGTTAATACCAGAGTG
AJ111D	CTTCATGGATGTGGACAACTAG

NT primers	
MS607A	CCTCTACAACCAGGTGCGACTGTG
MS607B	GCAGAGACAAGCCAGTAGGTATA

FB336R cosmid primers	
For1	CTGGGCATTCAATTCCTCAGG
Rev1	TACTAGGAAGAGGCATAAGG
6a8fa	GATCACTCATTGGCTCTCTCC
6a8fb	TACCTGACATTATAGCTTCC
6AA1	TGGAGCCCGGCAAATGCTTG
6AA2	TTGAGCTGGGTACAGGCAGC
6AA3	GAGGAGCGCTGACTGTGAGT
6AB1	CACCATTCTGCAGCCTGAGG
6AB2	TGCCTCAAGGTGGAAGCGAG

Methods

Specific reaction conditions and modifications to methods are described in the relevant chapters.

Tissue culture

All tissue culture manipulation was carried out in a category II laminar flow hood in a designated tissue culture laboratory. Cells were incubated in 5 % CO^2 at 37°C and high humidity in a New Brunswick CO^2 incubator (model CO27) using standard techniques. Fibroblast cell lines were grown in Dulbecco's Modified Eagle Media (Gibco BRL) supplemented with 10 % fetal calf serum (Gibco BRL) and lymphoblastoid cell lines were grown in RPMI 1640 medium (Gibco BRL) supplemented with 10 % fetal calf serum.

DNA extraction

I. Human genomic DNA from tissue culture cells

Cells were trypsinised or mechanically broken up to form a single cell suspension, and pelleted by centrifugation at 1100 rpm for 10 minutes. Cells were re-suspended in 1x PBS (sterile), counted, then re-pelleted and re-suspended in 500 µl 1x SSC (sterile) at a cell concentration of 10^7 cells/ml. Cells were lysed by addition of 500 µl lysis solution (100 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM EDTA, 1 % sarkosyl) and mixing until the solution became clear. Proteins were digested using proteinase K at a final concentration of 100 µg/ml and incubating at 55°C for six hours. The resulting solution was phenol/chloroform/isoamyl alcohol (25:24:1) extracted, and the aqueous phase removed, avoiding all interface layer. The DNA was ethanol precipitated and re-suspended in 1x TE. DNA concentration was determined using agarose gel electrophoresis against known amounts of marker DNA and other genomic DNAs of known concentration.

II. Propagation of *E. coli* and double stranded plasmid DNA

Plasmids were propagated in *E. coli* XL1Blue strain (Stratagene) in bacterial nutrient broth media containing ampicillin to 100 µg/ml and tetracycline to 25 µg/ml. DNA was obtained by an alkaline lysis procedure (Birnboim and Doly, 1979).

III. Single-stranded plasmid DNA from bacteria

The method from the pBluescriptII Exo/Mung DNA sequencing system protocol (Stratagene) was used, which involves the infection of XL1Blue culture at log growth phase (2.5 x 10^8 bacteria/ml) with VCSM13 helper phage (Stratagene) at a multiplicity of infection of 20:1 phage to cells. Incubation overnight allows

secretion of single-stranded DNA as packaged phage particles, which are precipitated from the medium, and single stranded DNA is then extracted.

Enzymatic manipulation of DNA

Enzymatic manipulation of DNA was carried out in the reaction buffer supplied with the enzyme according to the conditions recommended by the supplier unless stated otherwise.

DNA amplification

DNA was amplified using the Polymerase Chain Reaction (PCR) (Saiki *et al.*, 1988). PCR reactions contained (unless stated otherwise), 50 ng human genomic DNA or 50-100 ng plasmid DNA, 1 μM of each primer, 0.1 U/μl Taq polymerase (Advanced Biotechnologies), and 1x PCR buffer (kept as an 11x stock: 45 mM Tris-HCl pH 8.8, 11 mM $(NH_4)_2SO_4$, 4.5 mM $MgCl_2$, 6.7 mM β-mercaptoethanol, 4.4 mM EDTA, 1 mM dATP, 1 mM dCTP, 1 mM dGTP, 1 mM dTTP and 113 μg/ml BSA in a 10 μl reaction (Jeffreys *et al.*, 1988). Long-PCR reaction conditions (Barnes, 1994) were used for amplification of products greater than 2 kilobases, and contained 50 ng human genomic DNA, 1 μM of each primer, 1x PCR buffer, additional Tris-base (0.33 M) and 0.1 U/μl of Taq polymerase and cloned Pfu polymerase (Stratagene) in a 20:1 mix. Taq polymerase sometimes incorporates base mismatches into the new strand, which stalls the polymerisation. Pfu removes base mismatches, allowing Taq to resume extension of the new strand. Additional Tris in the reaction buffer increases the pH of the reaction at high temperatures, reducing the risk of template depurination. For long-PCR, denaturation was carried out at 94/95°C and extension times were increased to allow for synthesis of long amplicons (Cohen, 1994). All PCR reactions were carried out using a PTC200 DNA Engine™ (MJ Research, USA). Reactions were set up in a category II laminar flow hood, and PCR reagents and dedicated equipment were held separately from amplified products to prevent sample contamination. Zero DNA controls were always included. Details of specific temperatures and cycles are included in the relevant chapters.

Agarose gel electrophoresis

Agarose gel electrophoresis was carried out using horizontal submarine gels in 1x TAE (40 mM Tris-acetate, 1m M EDTA) or 0.5x TBE (45 mM Tris-borate, 1 mM EDTA) containing 0.5 μg/ml ethidium bromide. Electrophoresis tanks were manufactured in-house and power packs were supplied by Bio-Rad. DNA was visualised using a UV wand (Chromato-vue UVM-57, UVP Inc.) or a UV transilluminator (UVP High Performance transilluminator, UVP Life Sciences, Cambridge, UK.). The size of the DNA fragments to be resolved dictated the concentration and type of agarose used.

Preparative agarose gel electrophoresis

The band of interest was located in the gel, a slot cut behind it, and a narrow hole slightly wider than the well cut in front of the DNA with a clean scalpel blade. Pieces of dialysis membrane, prepared by boiling for ten minutes in 1x TE (10 mM Tris, 1 mM EDTA), were inserted into the cuts in the gel so that the membrane curled under the gel, and folded over the top of the gel, to prevent loss of DNA around the membrane. The gel was run at 3.33 V/cm for the required time, allowing the DNA to electroelute onto the membrane. Electroelution was monitored using a UV wand. With continuous application of the current, the membrane was then smoothly removed from the gel and placed into a microcentrifuge tube or a sterilin tube (Bibby Sterilin, UK), with a corner of the membrane trapped in the lid. Droplets of buffer containing the DNA fragment of interest were collected from the dialysis membrane by centrifugation at 4000 rpm for 4 minutes, in a benchtop Megafuge 1.0R (Heraeus Sepatech). The membrane was washed with 20-50 µl distilled water and centrifuged for a further 4 minutes. The eluate was then transferred to a fresh tube and the DNA recovered by ethanol precipitation.

Southern blot analysis and hybridisation

DNA was resolved on agarose gels and photographed. The position of the size marker bands were recorded by cutting small slots in the gel adjacent to the bands. The agarose gel was then depurinated in 0.25 M HCl for 2 x 7 minutes, alkali-denatured in 0.5 M NaOH for 2 x 15 min and neutralised in 0.5 M Tris-HCl, 3 M NaCl for 2 x 15 minutes. The DNA was transferred to a Hybond-N^{fp} (Amersham) nylon membrane by the capillary transfer method (Southern blotting; Southern, 1975) from the underside of the gel, using 20x SSC as the transfer buffer. The blotting process was carried out for a minimum of four hours but generally overnight. The position of the size marker bands were recorded on the membrane using a ball point pen corresponding to the slots previously marked on the gel. The membrane was washed in 6x SSC, dried and the DNA covalently linked to the membrane by exposure to 7 x 10^4 J/cm^2 of UV light in an Ultraviolet crosslinker (Amersham). Specific DNA sequences were detected by hybridisation to a radioactively labeled probe.

Double stranded DNA probe (5-10 ng) was labeled by the random primed labeling reaction (Feinberg and Vogelstein, 1983; 1984) which involves the use of randomly generated hexamers and DNA polymerase Klenow fragment to incorporate [α-^{32}P] dCTP into the DNA. The labeling reaction was stopped by the addition of 1.6 volumes of 'stop solution' (20 mM NaCl, 20 mM Tris-HCl pH 7.5, 2 mM EDTA, 0.25 % SDS), and the probe recovered from unincorporated deoxynucleotides by ethanol precipitation using 100 µg high molecular weight herring sperm DNA as a carrier.

Probes were dissolved in 0.5 ml distilled water and denatured by boiling for 3 minutes immediately prior to use. Southern blots were pre-hybridised for at least 30 minutes at 65°C in 'BLOTTO' (1.5x SSPE, 1 % SDS, 0.5 % non-fat dried milk, 6.1 % PEG 6-8000) for single copy probes, and in 7 % SDS, 0.5 M Na_2HPO_4 pH7.2, 1 mM EDTA (modified from Church and Gilbert, 1984) for repetitive probes. The pre-hybridisation solution was replaced with fresh solution before addition of the probe. Hybridisation was carried out at 65°C for 4-16 hours in a Maxi 14 hybridisation oven (Hybaid). After hybridisation the membrane was washed in at least 4 changes of wash solution, 0.2 x to 1x SSC/0.1 % SDS, depending on stringency required.

Single stranded synthetic oligonucleotide probes (30 ng) were labeled by phosphorylation of the 5′ termini with [γ-^{32}P]ATP by bacteriophage T4 polynucleotide kinase (Sambrook *et al.*, 1989) in buffer recommended by the manufacturer (Gibco BRL). Membranes were pre-hybridised for 30 minutes at 65°C in 0.05 % bovine serum albumin (BSA), 0.05 % Ficoll, 0.05 % Polyvinylpyrolidone (PVP), 0.5 % SDS, 0.1 % disodium pyrophosphate anhydrous ($Na_2H_2P_2O_7$), 5x SSC (modified from Denhardt, 1966). Hybridisation was carried out at 48°C overnight in a Mini 10 hybridisation oven (Hybaid). After hybridisation the membrane was washed in 4x SSC, 0.1 % SDS, 1 % disodium pyrophosphate for 4x 5 minutes.

The pattern of hybridisation was visualised by autoradiography using Fuji RX100 X-ray film, either at room temperature for very strong signals, or at –80°C with an intensifying screen.

Screening bacterial colonies by hybridisation

This was carried out using the method of Buluwela *et al.*, 1989. Colonies were grown overnight on Hybond-N^{fp}, lysed with 2x SSC/5 % SDS, and the DNA fixed to the nylon membrane by exposure to microwaves. Protein debris was removed from the membrane by washing in 5x SSC/0.1 % SDS. Hybridisation was then carried out as described above (Southern Blotting and Hybridisation).

Manual DNA sequencing

DNA sequencing from single stranded plasmid templates was carried out using the dideoxy chain-termination method (Sanger *et al.*, 1977) with a protocol devised by Stratagene (Sequenase™ protocol). Completed sequencing reactions were prepared for loading onto a gel by addition of formamide loading dye (95 % de-ionised formamide, 20 mM EDTA, 0.05 % xylene cyanol FF, 0.05 % bromophenol blue), and were heat denatured at 80°C for 10 minutes immediately prior to loading the gel. Samples were run on denaturing polyacrylamide gels containing 6 % acrylamide

(19:1 acrylamide:bisacrylamide, National Diagnostics), 7.67 M urea and a buffer gradient of 0.5x to 2.5x TBE for short reads or 0.5x to 1.5x TBE for long reads (Biggin *et al.*, 1983) in Sequi-Gen II Nucleic Acid Sequencing Cells (Bio-Rad). The gel was soaked in 10 % methanol, 10 % glacial acetic acid for approximately 15 minutes after completion to fix DNA in the gel and remove urea. The gel was transferred to 3MM paper (Whatman) and dried under vacuum with heat using a gel dryer (model 583, Bio-RAD) for two hours before autoradiography at room temperature using Fuji RX100 X-ray film.

Automated DNA sequencing

Sequencing was carried out using a PE Applied Biosytems Model 377 DNA Sequencing System, with the ABI PRISM BigDye™ Terminator Cycle Sequencing Ready Reaction Kit, in accordance with manufacturers instructions. Reactions were purified by ethanol precipitation and re-suspended in 2 µl 83 % de-ionised formamide, 8.3 mM EDTA, ready to be denatured for loading onto the gel. Gel running and analysis was carried out in the Protein and Nucleic Acid Laboratory (PNACL), University of Leicester.

Construction of telomere-anchored PCR libraries

Telomere-anchored PCR libraries were constructed using a method similar to that described in Royle *et al.*, 1992. High molecular weight DNA (25 µg) was digested with *Mbo*I, and fragments larger than 5 kb size selected by two rounds of preparative gel electrophoresis. Size selected fragments were ligated to a 100x molar excess of Sau3AI-linkers. Linkers were generated by phosphorylation of oligonucleotide SauL-B with T4 polynucleotide kinase and ATP, then annealing the SauL-A and SauL-B oligonucleotides, which form the top and bottom strands of the linker. The Sau3AI-linker has a 5′ overhang compatible with products of *Sau*3A/*Mbo*I digestion. Unligated linker molecules and linker dimers were removed by a further two rounds of preparative gel electrophoresis. Size selected DNA ligated to Sau3AI-linkers was amplified with cycle titration in parallel PCR reactions with primers TelB and SauL-A or TelC and SauL-A, and products detected by hybridisation with the telomere repeat probe. TelB anneals to the C-rich strand of the telomere repeat array, and amplification with TelB can indicate the presence of interstitial telomere-like repeat arrays. The optimal number of PCR cycles was then determined to be the one which gave a reasonable level of amplification using TelC but little or no amplification with TelB. The amplification was then scaled up using only TelC and SauL-A. PCR products were digested with restriction enzymes *Eco*RI and *Kpn*I. These restriction sites are present in the telomere primer and the linker primer respectively. The digested products were then size selected to remove the end fragments from the primer and the linker,

and cloned into the *Eco*RI and *Kpn*I sites of pBluescriptII SK+. Ligations were transformed into XL1Blue by electroporation, and plated out on bacterial agar containing x-gal (25 μg /ml), IPTG (25 μg/ml), ampicillin (100 μg/ml) and tetracycline (25 μg/ml). Plates were grown overnight at 37°C, and white colonies picked into 96 well microtitre plates containing 100 μl of a medium consisting of nine parts nutrient broth, one part 1x HMFM (3.6 mM K_2HPO_4, 1.3 mM K_2HPO_4, 2 mM sodium citrate, 1 mM $MgSO_4$, 4.4 % glycerol), ampicillin (100 μg/ml) and tetracycline (25 μg/ml) in each well. Plate position H12 was not inoculated in order to allow orientation of the filters after hybridisation, and position A1 contained vector without insert as a control for telomere repeat probe hybridisation. The microtitre plates were grown at 37°C for 24 hours, replicated onto nylon filters by 'hedgehog' (Coulson *et al.*, 1986) and screened by colony hybridisation.

Generating telomere repeat probes

The double-stranded telomere repeat probe Tel+ (TTAGGG) was generated in a PCR reaction with complimentary primer pair Tel1 and Tel2, without target DNA. Tel1 and Tel2 contain three repeat units and can anneal out-of-register, serving as a template for the generation of longer (TTAGGG)n products. Cycling primers at 96°C for 1min, 60°C for 1 min and 70°C for 1 min resulted in generation of double stranded products of varying lengths. Products of approximately 100 bp-1.3 kb were gel purified and ethanol precipitated. Double-stranded telomere variant repeat probes TelG (TGAGGG), TelJ (TTGGGG) and TelK (TCAGGG) were generated as above with primer pairs TelG/TelGcomp, TelJ/TelJcomp and TelK/TelKcomp. Products of approximately 100 bp-1.3 kb were gel purified and ethanol precipitated.

Construction of telomere-anchored PCR libraries using hybridisation selection

Small nylon filters bearing either mixed variant repeat sequences (TCAGGG), (TGAGGG) and (TTGGGG), or (TTAGGG) telomere repeat sequences only, were generated for filter hybridisation. Synthetic telomere repeat sequences were generated as described above. 1 μg of double-stranded Tel+ (TTAGGG) telomere repeat probe was denatured by treatment with alkali for 5 minutes (KOH, final concentration 100 mM), neutralised by adding 0.25 volumes of 1 M Tris-HCl pH 4.8, and a total of 0.1 μg DNA spotted onto a small (3 mm x 3 mm) nylon filter (Hybond-N^{fp}, Amersham). 1 μg of each of the double stranded variant repeat probes TelG, TelJ and TelK was pooled and treated in an identical manner, and a total of 0.1 μg DNA spotted onto nylon filter. When dried, filters were exposed to 7 x 10^4 J/cm^2 of UV light in an Ultraviolet crosslinker (Amersham) to covalently link the DNA to the filter. Filters were prehybridised in 1 ml phosphate/SDS buffer for 1 hour

(Church and Gilbert, 1984). Hybridisation was carried out in 100 µl of the same buffer in an Eppendorf tube.

Template DNA for hybridisation selection was prepared as follows. Human genomic DNA (25 µg) from one individual was digested to completion with *Mbo*I, and fragments larger than 5 kb size selected by two rounds of preparative gel electrophoresis. Size selected fragments were ligated to a 100x molar excess of Sau3AI-linkers. Unligated linker molecules and linker dimers were removed by a further two rounds of preparative gel electrophoresis. Approximately 50 ng of digested products were amplified in a PCR reaction with primers TelC and SauL-A. Cycling conditions were 96°C for 50 sec, 65°C for 40 sec and 70°C for 5 minutes, for 10 cycles. PCR products were recovered using a QIAquick PCR purification spin column (Qiagen Ltd, UK) in a 30 µl volume. DNA (15 µl) was denatured with alkali for 5 minutes (KOH, final concentration 25 mM), and neutralised by adding 0.2 volumes of 1 M Tris-HCl pH7.5. The resulting solution (21 µl) was added to 100µl hybridisation buffer containing the variant repeat filter and incubated overnight at 65°C. Paraffin oil was used to cover the reaction and prevent evaporation of hybridisation buffer. The variant filter was then removed, and the buffer containing unhybridised DNA (approximately 120 µl volume) was heat denatured and incubated overnight with the TTAGGG filter at 65°C. The filter was then thoroughly washed with 0.5x SSC/0.1 % SDS. PCR products bound to this filter were recovered by incubation of the filter in 100 µl of 10 mM KOH/0.01 % SDS for 5 minutes, shaking, at room temperature, followed by neutralisation in 100 µl of 0.5 M Tris-HCl (pH 7.5)/0.01 % SDS for 5 minutes, shaking, at room temperature. The two 100 µl washes were pooled, and DNA recovered in a final volume of 30 µl using QIAamp spin columns (Qiagen, UK), using the SauL-A primer at a final concentration of 1 µm as a carrier. A 1 µl volume of this fraction was then reamplified with linker primer SauL-A and the TelC telomere primer, and the optimum cycle number was determined by cycle titration and Southern blotting as described above. PCR products were digested with *Eco*RI and *Kpn*I and then cloned into pBluescriptII SK+ vector (Stratagene). Clones were arrayed in 96-well microtitre plates and replicated onto nylon filters. A schematic diagram of this strategy is shown in chapter 5, figure 5.1.

Computer sequence analysis

Computer-aided sequence analysis was carried out using the Genetics Computer group (GCG) sequence analysis software version 9.1 (Devereux *et al.*, 1984), running on a Silicon Graphics Inc. system using IRIX release 6.4. Analysis of automated sequencing data was carried out using Factura™ Release 1.2.0 (Applied

Biosystems) and Autoassembler Release 1.4.0 (Applied Biosystems) on Apple Macintosh computers.

Photography

Photographic records of ethidium bromide stained gels was carried out by visualisation of the products on a UV transilluminator (UVP High Performance transilluminator) and photography using a Polaroid MP-4 camera with Kodak negative film (T-Professional 4052) or a video system (UVP Life Sciences, Cambridge, UK). Negatives and autoradiographs were developed using reagents recommended by Kodak. Laser photocopies and laser/colour prints were generated by the Central Reprographic Unit (University of Leicester).

Chapter 3

ANALYSIS OF ORDERED ARRAY LIBRARIES GENERATED FROM PATIENTS WITH SUSPECTED TERMINAL DELETIONS

Summary

A telomere-anchored PCR approach was used successfully to isolate a terminal deletion breakpoint from an individual with a deletion at 7q32 (N. J. Royle, unpublished data). This chapter describes the use of this telomere-anchored PCR strategy to isolate telomere-junction clones from patients with suspected terminal deletions on 1p, 2q and 22q. Characterisation of clones from ordered array libraries resulted in the identification of a novel telomere repeat array from individual CB0001. The sequence adjacent to the telomere is a low copy subterminal repeat, which is unusually truncated by the telomere in this individual. The frequency of this polymorphic telomere is 2.4 % in the Caucasian population, but the chromosomal location has not been determined. A novel subterminal repeat found adjacent to a telomere on chromosomes 7, 11 and 17 was also isolated from individual CB0001. However, terminal deletion breakpoints were not isolated from telomere-junction clone libraries described in this chapter. Sequence data was compiled to generate a database that can be rapidly queried using local search tools, avoiding unnecessary investigation of previously isolated chromosome ends. The majority of telomere repeat arrays from normal chromosome ends contained variant repeats TGAGGG, TTGGGG and TCAGGG within the repeat array. Variant repeat probes were developed to detect clones containing these repeat types, and in conjunction with a panel of subterminal repeat probes, were used to eliminate a large number of clones containing normal chromosome ends prior to sequencing. A number of limitations with the general strategy may have resulted in failure to isolate terminal deletion breakpoints from the libraries described in this chapter. Telomere-anchored PCR amplification from only a small number of telomeres may result in the exclusion of some chromosome ends, including breakpoint fragments. Preferential cloning of small PCR products may also exclude terminal deletion breakpoints if the terminal *Mbo*I restriction site lies some distance from the breakpoint.

Introduction

Chromosomes with terminal deletions are thought to contribute to a number of different syndromes, including some cases of idiopathic mental retardation. Deletions of some chromosome ends may not give a distinct clinical phenotype, consequently terminal deletions may frequently go unidentified as an important cause of human genetic disease, particularly microdeletions of less than 1-2 Mb that cannot be detected by normal cytogenetic methods. One of the mechanisms by which terminal deletions may arise is via a chromosome break in the germline, or very early on in development, which is then healed and stabilised by *de novo* addition of a telomere, allowing the truncated chromosome to replicate and segregate normally. Evidence for this type of germline healing event has been found by study of six terminal deletions at the tip of 16p, resulting in the phenotype of α-thalassaemia (Flint *et al.*, 1994), a terminal deletion at 22q13.3 in patient NT (Wong *et al.*, 1997), causing mild mental retardation, and at 7q32, in a patient with holoprosencephaly (N. J. Royle, unpublished data). In all cases, the truncated chromosomes appear to have been healed by the direct addition of telomere repeats to non-telomeric sequence.

Cloning of telomeres has enabled the characterisation of sequences found adjacent to telomeres (Brown *et al.*, 1990; Cheng *et al.*, 1990; Cross *et al.*, 1989; de Lange *et al.*, 1990; Royle *et al.*, 1992). These regions are usually composed of several members of low copy repetitive sequence families which are often shared by a number of chromosome ends. A simple PCR-based method was developed to isolate short stretches of non-telomeric subterminal DNA adjacent to arrays of telomere repeats (Royle *et al.*, 1992). This telomere-anchored PCR method resulted in the isolation of telomere-junction clones, enabling detection and characterisation of telomere-associated repeats, and the pattern of variant repeats within the telomere repeat array. Most telomere-junction clones hybridised to multiple sequences in the human genome, representing sequence families present in most or all subterminal regions. One such family is represented by telomere-junction clone Tsk37 (Royle *et al.*, 1992). This subterminal repeat is present on many different chromosomes, and is related to a number of other subterminal sequences such as TelBam3.4, TelSau2.0 (Brown *et al.*, 1990), pTH2, pTH14 (de Lange *et al.*, 1990) and PGB4G7 (Cross *et al.*, 1990), as shown in figure 3.1a. There are approximately 200 copies of sequences related to this family in the human genome (Royle *et al.*, 1992) and different members can show between 2 and 20 % divergence (Cross *et al.*, 1990), suggesting that these sequences are evolving rapidly. The size and structure of this repeat family is not completely characterised, but it is complex and has internal repetition, with several arrays of tandem repeats (see figure 3.1a).

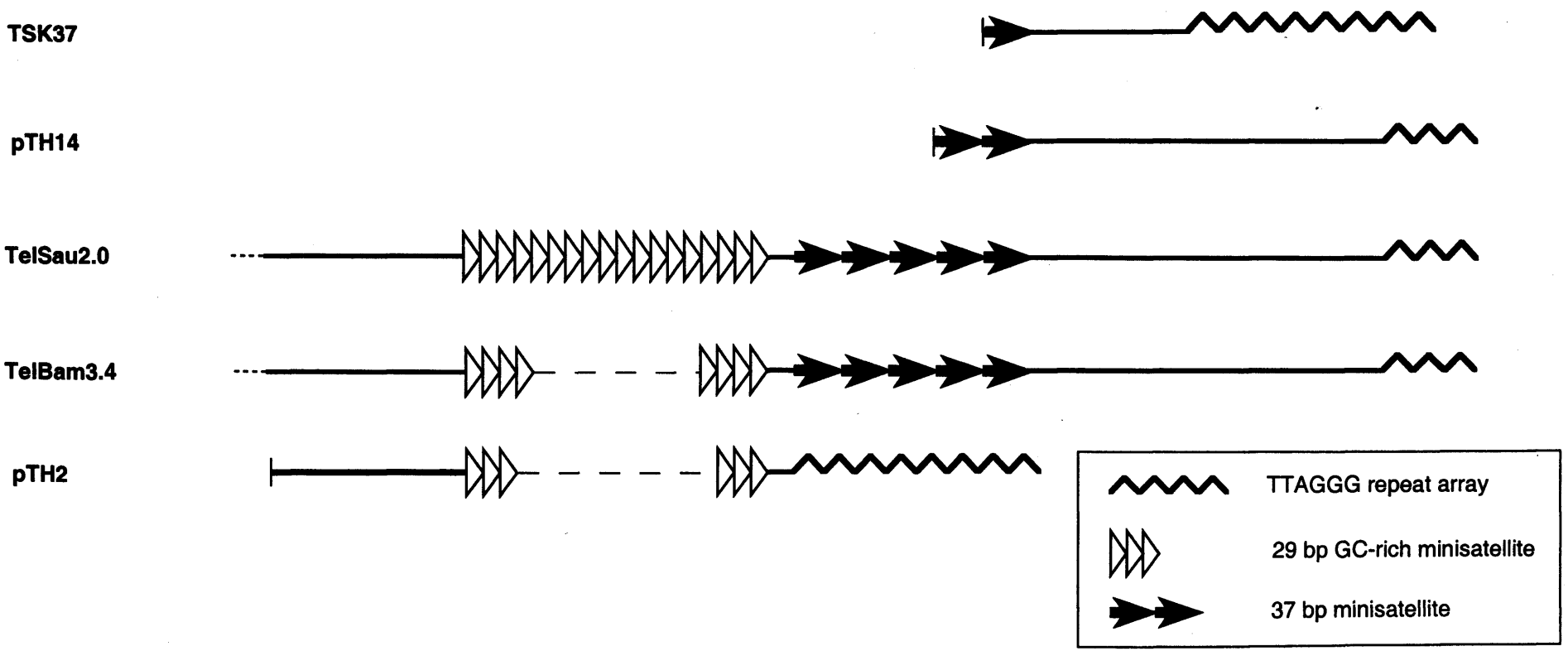


Figure 3.1a The relationship between different isolates of a human subterminal repeat family

The start of the cloned sequence is indicated by a vertical bar. Subterminal DNA not containing tandem repeats is represented by horizontal lines. Tandem repeat arrays are shown by arrows, and a dashed line has been introduced into these arrays to maximise alignment. The relationship between TelSau2.0 and TelBam3.4 is indicated from positions 1296 and 2313 respectively (Brown *et al.*, 1990). pTH2 and pTH14 have been previously described by de Lange *et al.* (1990). Figure from Royle (1995).

Nearest to the telomere repeat array there is a discontinuous array of five 37 bp repeats, and then a tandem array of 29 bp GC-rich repeats with variable copy number. Proximal to this is another tandem array of 61 bp repeats, also with variable copy number. The 29 and 61 bp repeats have a similar structure to GC-rich minisatellites located more proximally in the proterminal regions of human chromosomes (reviewed Royle, 1995). The subterminal repeat can be truncated at different positions before the start of the telomere repeat array. Different copies of the repeat show less similarity at the proximal ends, and some copies contain *Alu* elements and truncated L1-like sequences. Dispersed repeats such as *Alu* elements are common in subterminal GC-rich domains (Korenberg and Rykowski, 1988). Other telomere junctions isolated by telomere-anchored PCR included a repeat family found at the telomere-junctions of 7q and 12q, and the unique sequence found at the Xp/Yp pseudoautosomal telomere-junction (Tsk8), which does not cross-hybridise with subterminal repeats on other chromosomes (Brown *et al.*, 1990; Royle *et al.*, 1992). These chromosome ends may contain very diverged copies of other repeat families, or repeat families unique to these regions. Members of subterminal repeat families are located mainly at telomeres, but are also found near internal telomere-like repeats (Wells *et al.*, 1990).

The telomere-anchored PCR strategy was also used successfully to isolate telomere-adjacent sequences from the chimpanzee genome (Royle *et al.*, 1994). A chimpanzee subterminal satellite was isolated, composed of a 32 bp AT-rich repeat, which hybridised to all additional heterochromatic terminal bands not present in the human karyotype (approximately half of all chimpanzee chromosome ends), and two interstitial sites in the chimpanzee genome. This repeat probably represents an expanded satellite which has dispersed to many ends in the chimpanzee genome but has been lost from the human genome. This satellite is absent in orangutans, suggesting rapid evolution of terminal sequences in higher primates. Most human telomeres appear to have arisen since the human and chimpanzee divergence (Baird and Royle, 1997). Furthermore, telomeres appear to have been dynamic and relatively transient structures during the evolution of the great apes (Baird and Royle, 1997). Most chromosome ends are thought to undergo rapid turnover of terminal sequences, and mechanisms probably include terminal deletions, unequal exchange between subterminal repeats on non-homologous chromosomes, and exchange of subterminal sequence between homologous chromosomes due to shared sequences in these regions (Baird and Royle, 1997). Chromosome ends bearing unique or highly diverged subterminal sequence may have a reduced frequency of inter-chromosomal interaction.

The telomere-anchored PCR approach could potentially be applied to rapidly isolate breakpoint junctions from patients with terminal deletions by identifying a small region of non-terminal sequence at the breakpoint, adjacent to the newly synthesised telomere repeats. This method was used to generate ordered array libraries containing telomere-junction clones from three patients (FB336R, FB241 and NT) with suspected terminal deletions (N.J. Royle, personal communication). Patients FB336R and FB241 have a phenotype of mild holoprosencephaly (HPE), a developmental field defect involving the brain and face, including mental retardation and mild facial dysmorphia. One of the genes involved in the HPE phenotype has been localised to 7q36 (Frezal and Schinzel, 1991). The phenotypes in patients FB336R and FB241 were due to cytogenetically visible *de novo* deletions of 7q32-qter and 7q34-qter respectively (Gurrieri *et al.*, 1993). Patient NT has mild mental retardation and speech difficulties; overlapping features of the cytogenetically visible 22q13.3 deletion syndrome. The phenotype in NT results from a *de novo* microdeletion, and the breakpoint was localised to 22q13.3 by polymorphic marker analysis (Flint *et al.*, 1995).

The terminal-deletion breakpoint was identified in an ordered array library containing telomere-junction clones from patient FB336R. The clone containing the breakpoint had 304 bp of chromosome 7 unique sequence, adjacent to an array of TTAGGG repeats. PCR amplification with a primer from the telomere-adjacent sequence and a telomere primer, followed by Southern hybridisation, resulted in detection of a smear of hybridising products, characteristic of telomere amplification, in patient FB3336R but not in 79 other unrelated individuals. A cosmid clone containing the breakpoint sequence was localised by FISH on metaphase chromosome spreads to an internal location (7q32) in normal individuals, and on the patient's normal chromosome 7. The evidence is consistent with healing of the terminal-deletion breakpoint by direct addition of TTAGGG repeats, possibly by telomerase. Analysis of the FB241 and NT ordered array libraries failed to identify terminal deletion breakpoints (N.J Royle, personal communication). However, a low frequency polymorphic telomere repeat array was identified in patient NT, which truncated a subterminal repeat array not normally associated with telomere repeats (Coleman *et al.*, 1999). The breakpoint in NT was later cloned (Wong *et al.*, 1997), and located within minisatellite locus MS607 (D22S163; Armour, 1990), resulting in a 130 kb deletion. This breakpoint appears to have been healed by the direct addition of telomere repeats. The acrosin (ACR) gene maps within the deletion, located about 70 kb from the telomere. ACR encodes a serine protease present in the acrosome of the sperm head (Klemm *et al.*, 1991). It is unlikely that deletion of this gene contributes to the phenotype in NT,

but it is possible that other genes normally present in the deleted region have yet to be mapped (Wong *et al.*, 1997).

Telomeres contain consensus TTAGGG repeats, however, variant repeats such as TTGGGG and TGAGGG are often present (Allshire *et al.*, 1988) and tend to be clustered at the proximal end of the telomere repeat array. These variants may arise during replication, and persist because they are not subject to degradation, or the action of telomerase (Allshire *et al.*, 1988; Guerrini *et al.*, 1993). It is thought that telomeres containing variant repeats have been present at the chromosome end for some time in order to accumulate these mutations. Maps of the interspersion patterns of these variant repeats at some telomeres in many individuals have been generated by a PCR-based method called Telomere Variant Repeat (TVR) mapping (Baird *et al.*, 1995). The distribution of telomere and variant repeats at the proximal ends of the Xp/Yp pseudoautosomal telomere (Baird *et al.*, 1995), the 16q telomere (Coleman *et al.*, 1999) and the 7q and 12q telomeres (Baird *et al.*, 1999; Coleman, 1998) is highly variable, suggesting a high mutation rate in these regions, and this is likely to be a common feature of all telomeres. In the Xp/Yp telomere most of the variation stops at around 120 repeat units into the array (Baird, 1996). No distal limit of variation has yet been identified at the 7q and 12q telomeres, and TGAGGG repeats extend at least 200 repeats into the 12q array (Baird *et al.*, 1999). Southern blot analysis identified a distal limit of approximately 1.9 kb for variant repeats TGAGGG and TTGGGG in most telomeres, distal to this probably only TTAGGG repeats exist (Allshire *et al.*, 1989).

The terminal deletions at 16p13.3, 7q32 (FB336R) and 22q13.3 (NT) appear to have been healed by the direct addition of telomere repeats to non-telomeric sequence. There is no subterminal sequence present adjacent to these breakpoints, and the telomere repeat arrays contain TTAGGG repeats only. It is likely that telomere repeat addition was mediated by the enzyme telomerase. Evidence for chromosome healing by another mechanism, called telomere capture, has been found in a patient with a terminal deletion at 16p13.3 and α-thalassaemia. The truncated chromosome has been healed by capture of an existing telomere via an illegitimate recombination event between two related *Alu* elements (Flint *et al.*, 1996). Subterminal sequence was transferred to the breakpoint with the existing telomere, which probably contained variant repeat types. The presence or absence of subterminal sequence adjacent to the telomere, and presence or absence of variant repeat types in the telomere provides a means by which telomere capture and *de novo* telomere addition may be distinguished. However, the frequency at which chromosome healing by *de novo* telomere addition occurs is unknown, and

it is not clear whether this is a common mechanism by which broken chromosomes are healed, or a relatively rare event.

Aims of work described in this chapter

The telomere-anchored PCR-based method was used successfully to isolate the terminal-deletion breakpoint in patient FB336R. The main aim of the work described in this chapter was to use this simple method to isolate terminal-deletion breakpoints from a number of other patients that may have been healed by direct telomere addition, possibly by the action of telomerase. A further aim was to improve the efficiency of the screening step used for preliminary analysis of ordered array libraries containing telomere-junction clones. A more rigorous screening procedure would reduce the number of candidate breakpoint clones to be sequenced, making library analysis less time consuming. However, it should be noted that identification of a breakpoint is dependent on two assumptions; that the newly synthesised telomere contains only TTAGGG repeats without variants, and that no subterminal sequence remains adjacent to the telomere.

Results

3.1 Strategy for constructing ordered array libraries

Ordered array libraries were generated from patient genomic DNA as described in Chapter 2 (Materials & Methods - Construction of telomere-anchored PCR libraries). The general strategy is outlined in figure 3.1. Tandemly repeated DNA is deficient in many restriction enzyme sites, and telomeres do not contain *Mbo*I recognition sites. Consequently, digestion of high molecular weight genomic DNA with *Mbo*I generates a population of larger fragments mainly representing telomeres and other repeated sequences. Agarose gel size selection of fragments larger than 5 kb gives a 50-100 fold enrichment for large *Mbo*I fragments including telomeres. Ligation of Sau3AI-linkers to the size fractionated products allows telomere-anchored PCR amplification with the SauL-A linker primer and the TelC telomere primer, which anneals to repeats on the G-rich strand of the telomere. The PCR step enriches for the longer fragments containing telomere repeat arrays by generating short amplicons which can be cloned. Chromosome ends with short regions of telomere-adjacent sequence amplify very efficiently. Directional cloning into the pBluescriptII SK+ vector using *Kpn*I and *Eco*RI, that cut in restriction sites present in the SauL-A and TelC primers respectively, leads to cloning of only amplicons with a linker on one end and a telomere primer site on the other end. A small number of rare fragments with internal *Kpn*I or *Eco*RI restriction sites may also be cloned. Interstitial telomere-like repeat arrays may be cloned, although the majority of these regions appear to be less than 1 kb (Flint *et al.*, 1997) and are therefore mainly excluded during the size fractionation step. Residual contaminating arrays may be present due to the limitations of gel size fractionation but are under-represented. Amplification with primer TelB in parallel to amplification with TelC can indicate the presence of interstitial telomere-like repeat arrays. TelB anneals to the C-rich strand of the telomere repeat array and primes in a 5′ to 3′ direction. PCR cycle number is carefully controlled to further minimise the number of products arising from amplification of interstitial telomere-like repeat arrays. Cloned interstitial telomere-like repeats are not necessarily distinguishable from cloned telomere repeat arrays, although very degenerate repeats are a feature of these regions (Wells *et al.*, 1990). Replicate nylon filters from ordered array libraries were screened with radioactively labeled Tel+ telomere repeat probe (TTAGGG) to detect clones positive for telomere repeats. Following Tel+ hybridisation, filters were screened with a number of radioactively labeled subterminal probes consisting of subterminal repeat sequences Tsk37, Tsk46, Tsk8 (Xp/Yp pseudoautosomal telomere-junction) and E-F (7q and 12q telomere-junction; Royle *et al.*,1992). These subterminal repeat

Figure 3.1 Generation of ordered array libraries by telomere-anchored PCR

Telomere repeats are represented by blue arrows, Sau3AI-linkers by red bars. Green ovals and striped boxes represent other tandemly repeated regions not cut by *Mbo*I digestion. This method was adapted from the telomere-anchored PCR method described by Royle *et al.*, 1992.

1 High molecular weight genomic DNA digested with *Mbo*I.

2 Fragments larger than 5 kb size selected by preparative gel electrophoresis, Sau3AI linkers ligated onto the overhangs generated by *Mbo*I.

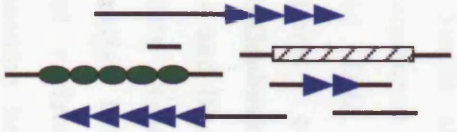
3 PCR amplification carried out using a linker primer and a telomere repeat primer, with cycle number carefully controlled to minimise products arising from interstitial telomere-like repeat arrays.

4 PCR products digested with *Kpn*I and *Eco*RI. Restriction sites are present in the linker primer and telomere primer respectively.

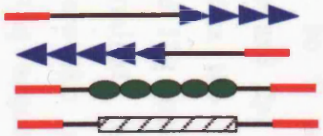
5 Digestion products directionally cloned into pBluescriptII SK+ vector and transformed into XL1Blue. Recombinant colonies picked into 96-well plates.

6 Colonies replicated onto nylon filters.

7-9 Filters screened with the telomere repeat probe and subterminal probes. Clones positive for telomere repeats but not subterminal probes sequenced.

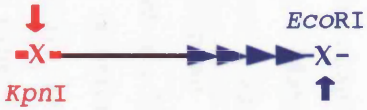


1 Digestion of high molecular weight genomic DNA with *Mbo*I



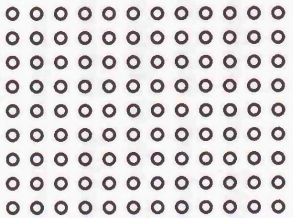
2 Size fractionation above 5kb and ligation of Sau3A-linkers

3 PCR amplification with linker primer and telomere primer

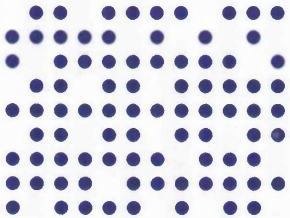


4 Digestion with *Kpn*I and *Eco*RI to target restriction sites in primers

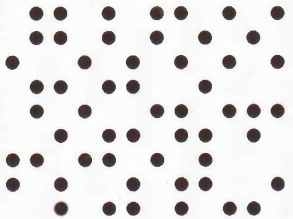
5 Ligate into pBluescriptII SK+, transform into XL1Blue, and pick recombinant colonies into 96-well plates



6 Replicate plates onto nylon filters

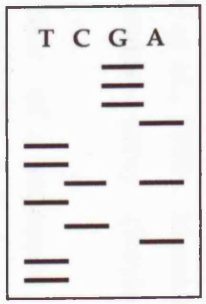


7 Screen filters with telomere repeat probe



8 Screen filters with subterminal probes

9 Sequence clones positive with telomere repeat probe but negative for subterminal probes



Chapter 3. Figure 3.1

regions, and the primers used to generate subterminal probes are shown in figure 3.2.

Tsk37 is a member of the family of telomere-associated sequences TelBam3.4, TelSau2.0 (Brown *et al.*, 1990), pTH2, and pTH14 (de Lange *et al.*, 1990). Members can be variably truncated before the start of the TTAGGG repeat array (Royle *et al.*, 1992).

Tsk46 (EMBL/Genbank accession no. X60164) is a member of a sequence family with at least one copy present near the terminus of chromosome 7. Some members of this family appear to be related to the Tsk37 repeat family.

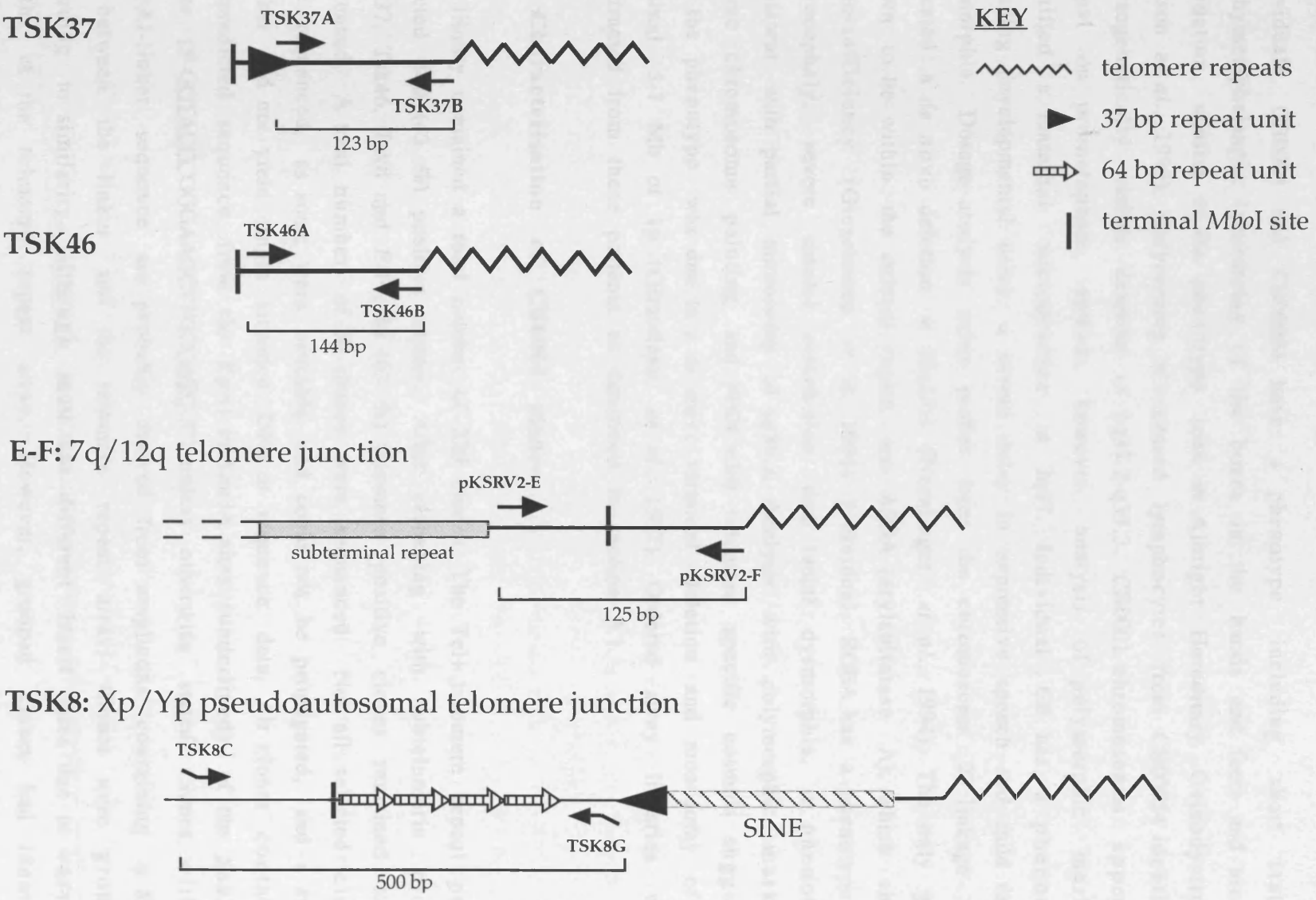
Tsk8 (Xp/Yp pseudoautosomal telomere-junction). The terminal 1 kb of the pseudoautosomal telomere junction contains a unique monomorphic G-rich minisatellite composed of four 64 bp repeat units, and a truncated copy of a short interspersed nuclear element (SINE) located 14 bp from the start of the telomere repeat array, orientated away from the telomere. The 500 bp Tsk8 probe encompasses the minisatellite but does not contain the SINE.

7q/12q (E-F). The terminal 125 bp of 7q and 12q show 95 % sequence similarity and appear to be distinct from other chromosome ends. This 125 bp region also shows 86 % sequence identity to two interstitial telomere-like repeat arrays located at 16p and 2p, but is only found at a telomere-junction on 7q and 12q. The E-F probe identifies the 125 bp region immediately adjacent to these telomeres. The 512 bp at 7q and 12q proximal to the 125 bp telomere-junction sequence shows homology to subterminal repeats found on many chromosome ends, and the interstitial telomere-like repeat arrays at 16p and 2p. Proximal to this 512 bp region, the 7q and 12q sequences show homology to a subterminal sequence family that is uncommon in the human genome (Baird *et al.*, 1999; Coleman *et al.*, 1999).

Screening ordered array libraries with the subterminal probes described above eliminated a number of telomere positive clones containing normal chromosome ends. Single stranded DNA from remaining telomere positive clones was extracted and sequenced as described in chapter 2 (Materials and Methods - Single-stranded plasmid DNA from bacteria, Manual DNA sequencing) to identify candidate breakpoint clones. Sequence identity searches were carried out at the EMBL/Genbank database (European Bioinformatics Institute, Cambridge, UK) using FASTA and BLAST (Altschul *et al.*, 1990) search programs. Criteria used to define a candidate clone were: a) telomere-adjacent sequence without homology to previously isolated subterminal sequences, and b) an array of TTAGGG repeats without variants. This is the expected structure of a breakpoint fragment from a chromosome healed by addition of a new telomere.

Figure 3.2 Structure of telomere-junction clones and location of primers used to make subterminal probes

TSK37 is a member of a family of telomere-associated sequences. This isolate had 123 bp of non-telomeric DNA containing one 37 bp repeat unit from a minisatellite found in this region. Probe Tsk37 (123 bp) was generated using primers TSK37A and TSK37B.
TSK46 is a member of another subterminal repeat family. This isolate had 149 bp of non-telomeric-adjacent DNA. Probe Tsk46 (144 bp) was generated using primers TSK46A and TSK46B.
E-F identifies a short region (125bp) found at the telomere-junctions of 7q and 12q only. The more proximal subterminal sequence at 7q and 12q show homology to other subterminal regions. The E-F probe (120 bp) was generated using primers pKSRV2-E and pKSRV-2F.
TSK8 identifies the Xp and Yp pseudoautosomal telomere-junction. TSK8 was generated using primers TSK8C and TSK8G, both of which have a non-complementary 5′ tail. The TSK8 probe is 500 bp and encompasses the four 64 bp repeat units of the minisatellite found in this region, but does not include the SINE.



Chapter 3. Figure 3.2

3.2 Construction of ordered array libraries from patients with suspected terminal deletions

Individuals CB0001 and CB00054 have a phenotype including short stature, brachymetaphalangia (shortening of the bones in the hands and feet) and mental retardation, similar to the phenotype seen in Albright Hereditary Osteodystrophy (Wilson *et al.*, 1995). Karyotyping of cultured lymphocytes from CB00054 identified a cytogenetically visible deletion of 2q37.2-q37.3. CB0001 chromosomes appeared normal on prometaphase spreads, however, analysis of polymorphic markers identified a maternal microdeletion at 2q37. Individual CB has a phenotype including developmental delay, a severe delay in expressive speech and mild facial dysmorphia. Dosage analysis using probes from the chromosome 22 linkage map indicated a *de novo* deletion at 22q13.3 (Nesslinger *et al.*, 1994). The only gene known to lie within the deleted region was ARSA (arylsulfatase A), which shows haplo-sufficiency (Gieselmann *et al.*, 1991). Individual ROBA has a phenotype of microcephaly, severe mental retardation and facial dysmorphia, a phenotype consistent with partial monosomy of 1p36.3. Analysis with polymorphic markers, reverse chromosome painting and FISH with telomere specific cosmids suggested that the phenotype was due to a *de novo* terminal deletion and monosomy of the terminal 5-7 Mb of 1p (Giraudeau *et al.*, 1997). Ordered array libraries were constructed from these patients as described in section 3.1.

3.3 Characterisation of CB0001 clones

The library contained a total number of 329 clones. The Tel+ telomere repeat probe detected 140 (43 %) positive clones. After screening with subtelomeric probes (Tsk37, Tsk46, Tsk8 and E-F), 88 (63 %) telomere positive clones remained to be sequenced. A total number of 50 clones were sequenced. Not all selected clones were sequenced, as some were unstable and could not be propagated, and a small number did not yield single stranded DNA or sequence data. All clones contained the predicted sequence from the *Kpn*I to *Sau*3A sites (underlined) of the Sau3AI-linker (5'-GGTACCCGGGAAGCTTGGGATC-3') unless otherwise stated. Clones without Sau3AI-linker sequence are probably derived from amplicons containing a *Kpn*I site between the linker and the telomere repeat array. Clones were grouped according to similarity, although most had different insert sizes due to varying lengths of the telomere repeat array. However, grouped clones had identical telomere-adjacent sequence and telomere position with respect to the flanking DNA (unless otherwise stated). Clones were sequenced from the *Sau*3A site in the linker to the *Eco*RI cloning site in the vector. The entire inserts of some clones were not sequenced if telomere repeat arrays were evident on the sequencing gel.

Sequence from large inserts was obtained by running sequencing gels for extended periods to generate long reads. A summary of clones sequenced from the CB0001 library is shown in table 3.1.

Group 1. Five clones contained TTAGGG and variant repeats only.

Group 2. Eighteen clones had 28 bp of identical telomere-adjacent sequence with 100 % identity to the telomere-junction of the PGSEGB telomere-associated sequence (EMBL/GenBank accession no. X56277; Cross *et al.*, 1990). The telomere repeat arrays contained variant repeats including TGAGGG and TTGGGG, and showed differences between clones within this group. Analysis of the TTAGGG and variant repeat types identified at least ten different interspersion patterns (shown in figure 3.3), suggesting that these clones represent at least ten different isolates of the PGSEGB sequence family adjacent to a telomere repeat array. These isolates may be allelic, or originate from different chromosome ends.

Group 3. Six clones had 49 bp of identical telomere-adjacent sequence with sequence similarity to ESTs found in subterminal regions of 22q (96 %; AC002055), 4q/10q (94 %; AF017467 and U74496) and 4ptel (90 %; Z95704; Baxendale *et al.*, 1993), shown in figure 3.4. However, these sequences are not normally located adjacent to telomere repeats. Consequently these clones were thought to represent a chromosome end polymorphic for presence of a telomere at this location. Investigation of this polymorphic telomere is described in detail in section 3.4. These clones had identical arrays of telomere repeats containing variants including TCAGGG, TGAGGG, TTGGGG and other variant repeats such as TAGGG.

Group 4. Eight clones had a *Kpn*I (GGTACC) start site and 13 bp of identical telomere-adjacent sequence. This region was too short to identify significant matches in the EMBL/GenBank database. The telomere repeat arrays all contained TTGGGG and other variant repeats, but displayed at least three different repeat interspersion patterns, representing different isolates or alleles of this telomere-junction.

Group 5. Four clones had a *Kpn*I start site and 129 bp of identical telomere-adjacent sequence. This sequence showed 100 % identity to a region of the PGB4G7 subterminal repeat (accession no. X56278; Cross *et al.*, 1990), which localises to 8p, 21p and 21q in different individuals. The four telomere repeat arrays contained variants TCAGGG and TCGGGG. Three different repeat interspersion patterns were identified among the four clones, representing different alleles or chromosomal isolates of this telomere-junction sequence.

Group 6. Two identical clones had 92 bp of identical telomere-adjacent flanking sequence, with 90 % sequence identity to the LTR13 (Long Terminal Repeat 13) consensus sequence (AL008730), and telomere repeat arrays containing TCAGGG variant repeats.

Table 3.1 Summary of clones sequenced from the CB0001 library

Clone names are shown in the first column. Clones are grouped according to similarity either in the telomere-adjacent DNA or telomere repeat array, if adjacent DNA is absent. The number of clones in each group, and the start site, *Sau*3A (GATC) or *Kpn*I (GGTACC), are shown in columns two and three. The length of telomere-adjacent sequence in each group of clones is shown in basepairs. Results of sequence homology searches using BLAST and FASTA search tools at the EMBL/Genbank database are shown in the fifth column. Presence or absence of telomere repeats, and TGAGGG, TTGGGG, TCAGGG or other variant repeats is shown in columns six and seven. The number of interspersion patterns within the telomere repeat arrays of each group of clones is indicated in column eight. The far right-hand column indicates whether a clone has been previously isolated and from which ordered array library.

Table 3.1 Summary of clones sequenced from the CB0001 library

Clone name	No. of clones in group	Start site	Telomere-adjacent sequence	Sequence identity FASTA/BLAST database search	Telomere repeats present	Variant repeats present	No. of repeat interspersion patterns in group	Previously identified?
GROUP 1	5	*Sau* 3A	none	–	yes	TCAGGG, TGAGGG other variants	–	–
GROUP 2	18	*Sau* 3A	28 bp	100 % PGSEGB subterminal repeat (X56277)	yes	TGAGGG, TTGGGG other variants	> 10	NT
GROUP 3	6	*Sau* 3A	49 bp	96 % 22q subterminal EST (AC002055) 94 %10q/4q subterminal EST (AF017467/U74496) 90 % 4ptel EST (Z95704)	yes	TCAGGG, TGAGGG, TTGGGG other variants	> 4	no
GROUP 4	8	*Kpn* I	13 bp	–	yes	TTGGGG other variants	> 3	NT
GROUP 5	4	*Kpn* I	129 bp	PGB4G7 subterminal repeat (X56278)	yes	TCAGGG other variants	> 3	NT
GROUP 6	2	Sau3A	92 bp	90 % LTR-13 consensus (AL008730)	yes	TCAGGG		FB336R
GROUP 7	2	*Sau* 3A	> 200 bp	98 % human satellite II (X72623)	no	–		NT FB336R
GROUP 8	3	*Sau* 3A	> 200 bp	no significant matches	no	–		no
212c		*Sau* 3A	630 bp	96 % Xp/Yp pseudoautosomal telomere junction (Tsk8)	yes	TTGGGG		no
24c		*Sau* 3A	327 bp	no significant matches	yes	TTGGGG		no

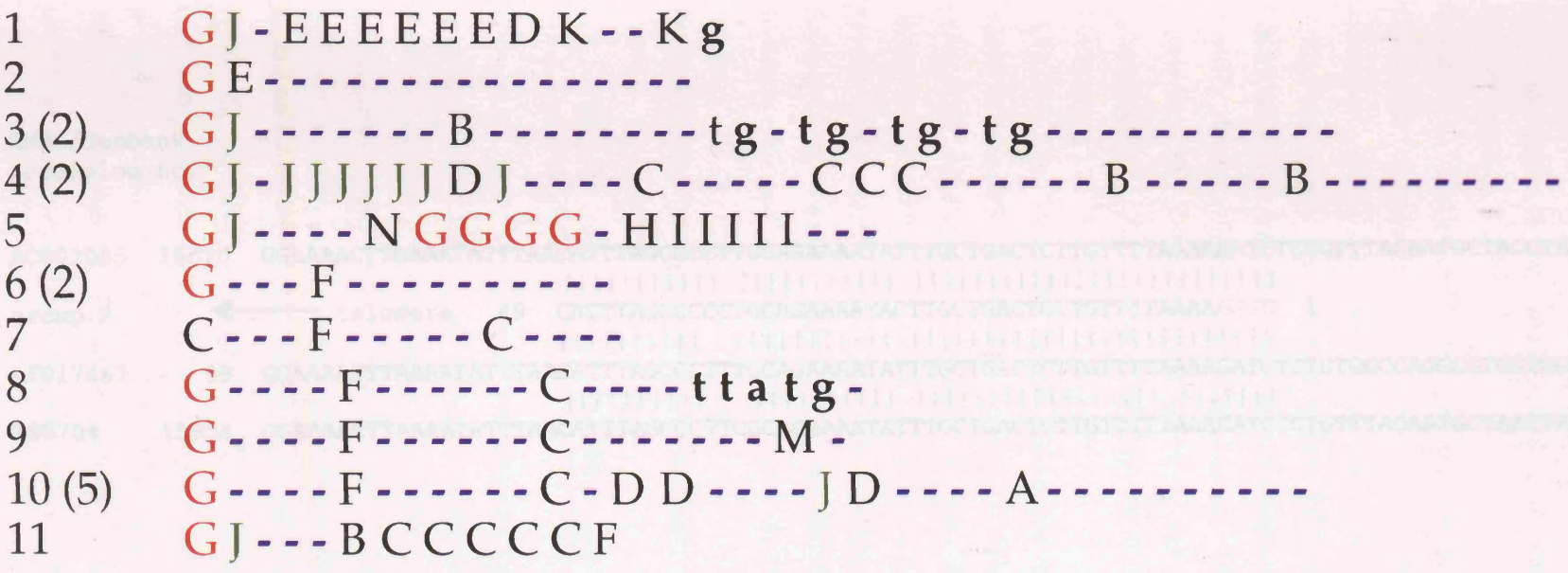


Figure 3.3 Telomere repeat interspersion patterns of group 1 clones in the CB0001 library

Repeat types are designated by the following characters (lower case indicates single base pairs):

-	ttaggg	A	ttagg	D	ttagggg	H	cgaggg	M	tagggg
G	tgaggg	B	tttggg	E	ttggggg	K	ttagagg	N	tttaggg
J	ttgggg	C	taggg	F	ttagcg	I	ctaggg		

Alignment of the telomere repeat arrays indicates that there are at least ten different repeat patterns within group 1 clones, representing different isolates of the PGSEGB telomere-associated sequence. Repeat arrays 6 and 7 are similar, differing by only two repeat units. These repeat arrays may have arisen from the same chromosome end, with the small variations arising during PCR amplification. Numbers in brackets indicate the number of clones in each pattern group, if more than one.

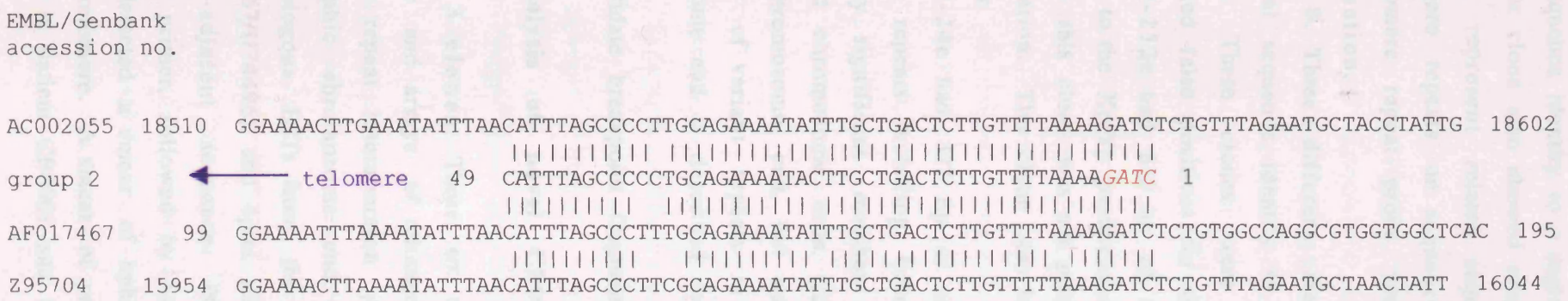


Figure 3.4 Alignment of the consensus sequence from CB0001 group 3 clones with other subterminal regions

Group 3 clones had 49 bp of telomere adjacent DNA and an array of telomere repeats with variants. The group 3 consensus is shown in the reverse orientation, with the direction of the telomere indicated by the blue arrow. The *Sau*3A cloning site is shown in red. The telomere-adjacent sequence shows homology to subterminal regions found on a number of different chromosomes – 96 % sequence identity to the 22q subterminal region (AC002055), 94 % sequence identity to the 4q/10q subterminal region (AF017467/U74496), and 90 % sequence identity to the 4p subterminal region (Z9570). These subterminal sequences are not normally found adjacent to telomere repeats. Vertical lines represent base matches between sequences. The orientations of AC002055, AF017467 and Z95704 are unknown with respect to the telomere.

Group 7. Two similar clones had telomere-adjacent sequence showing 87 % sequence identity to each other over a 128 bp region. One of these clones showed 98 % sequence identity to a region of human satellite 2 (X72623; Jeanpierre, 1994). The other clone also showed homology to this satellite 2 sequence, and these clones probably represent related sequences. However, these clones showed no evidence of telomere repeats on sequencing, and were only weakly hybridising with the Tel+ telomere repeat probe, and probably represent false positives due to cross-hybridisation.

Group 8. Three different clones had no telomere repeats and did not show significant sequence identity to each other or to sequences in the EMBL/GenBank database. These clones were not investigated further, and probably also represented false positives for the Tel+ telomere repeat probe.

CB0001-212c had 630 bp of telomere-adjacent sequence with 96 % sequence identity to the Xp/Yp pseudoautosomal telomere junction Tsk8 (Royle *et al.*, 1992). However this clone was not positive for the Tsk8 probe, possibly due to failure of hybridisation. This clone also had an array of telomere repeats including TTGGGG variants.

CB0001-24c had 327 bp of telomere-adjacent sequence, and an array of 20 telomere repeats including three TTGGGG variants. The flanking sequence did not show any significant matches in the EMBL/GenBank database or to previously identified chromosome ends. Consequently, this clone was thought to represent a novel chromosome end, but was excluded as a candidate breakpoint due to the presence of variant repeats in the telomere repeat array. Investigation of this chromosome end is described in the following section.

No candidate breakpoint fragments were isolated from the CB0001 library.

3.4 Analysis of novel CB0001 telomere-junction clones

Group 3 clones. These six clones had 49 bp of identical telomere-adjacent sequence and arrays of telomere repeats containing variants, with at least four different repeat interspersion patterns. These clones were thought to represent a polymorphic chromosome end, due to the lack of telomere repeats at this location in homologous ESTs from the subterminal regions of 22q (AC002055), 4q/10q (AF017467/U74496) and 4ptel (Z95704). Primer CB0001-14d was designed from the telomere-adjacent sequence. PCR amplification with this primer and the TelC telomere primer, followed by Southern hybridisation using primer CB0001-14d as a probe, detected a smear of hybridising products characteristic of amplification from a telomere. A smear of products with an underlying banding pattern was observed in patient CB0001 and father DW but not in mother BW (data not shown).

The underlying banding pattern probably arises due to the presence of variant and degenerate repeat blocks within the telomere repeat array, that do not amplify with the TelC telomere primer, generating gaps in the smear of hybridising products.

PCR amplification with primers CB0001-14d and TelC resulted in detection of a telomere repeat array in only CB0001 and 1 of 81 unrelated CEPH individuals (1331-01). The banding patterns in patient CB0001 and CEPH individual 133101 showed some differences, possibly due to expansion of blocks of variant repeats within the array, shown in figure 3.5. Banding patterns between unrelated individuals are expected to differ if telomeres are evolving rapidly. The presence of a telomere repeat array at this location in only 2 of 82 unrelated individuals tested suggests that these telomere-junction clones (CB0001 group 3) represent a low frequency polymorphic telomere, present in 2.4 % of the Caucasian population. The chromosomal location of this polymorphic telomere was not determined due to its low level of frequency in the Caucasian population. It is possible that these clones represent an internal telomere-like repeat array rather than a true telomere. However, most internal repeat arrays are less than 1 kb (Flint *et al.*, 1997). The smear of amplified products from this telomere-adjacent region extends up to 2 kb in the two individuals tested, suggesting that this is a terminal repeat array. The presence or absence of this polymorphic telomere in other populations was not investigated.

Clone CB0001-24c. This clone had 327 bp of telomere-adjacent sequence, and an array of 20 repeats including three TTGGGG variant repeats. The telomere-adjacent sequence did not show any significant matches in the EMBL/GenBank database, but the presence of variant repeats excluded this clone as a candidate breakpoint. Consequently, this clone was thought to represent a novel chromosome end. Primers CB0001-24cA and CB0001-24cB were designed to amplify from the telomere-adjacent sequence of this clone.

PCR amplification was carried out using primer CB0001-24cA and the TelC telomere primer, followed by Southern hybridisation with probe CB000124c (generated in a PCR reaction with primers CB0001-24cA and CB0001-24cB). A smear of hybridising products characteristic of telomere amplification was detected, with an underlying discrete banding pattern, in the patient CB0001 (KW), father DW, mother BW and all other individuals tested (human placental DNAs P50 and P53, and 16 CEPH individuals). Amplification from a number of these individuals is shown in figure 3.6. The CB0001-24c sequence is found adjacent to a telomere in most or all individuals. Small variations in the banding patterns in unrelated individuals are

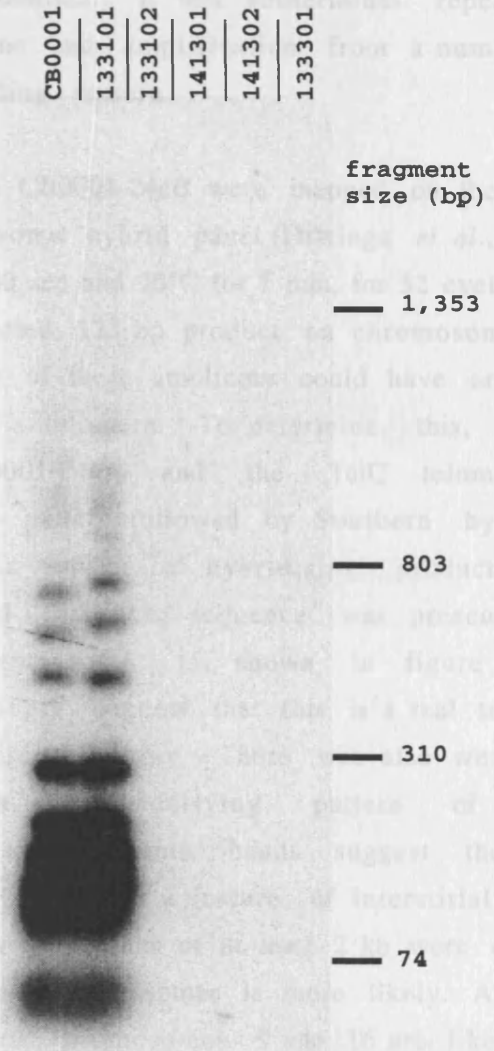


Figure 3.5 Banding patterns in the CB0001 polymorphic telomere in unrelated individuals CB0001 and 1331-01

Amplification was carried out using primers CB0001-14d and TelC, and products detected using the CB0001-14d oligonucleotide as a probe. Individuals tested are indicated above each lane. Products were detected in CB0001 and CEPH individual 1331-01 only. The banding patterns in these individuals show some differences which may be due to expansion of blocks of like-repeats in the telomere repeat array in individual 1331-01.

evident (see figure 3.6), and are probably due to different interspersion patterns of variant repeat types. For example, the banding patterns in individuals P50 and P53 are similar but not identical. If this subterminal repeat family is present at more than one chromosome end, amplification from a number of telomeres would generate a complex banding pattern.

Primers CB0001-24cA and CB0001-24cB were mapped on the NIGMS human/rodent somatic cell monochromosome hybrid panel (Drwinga *et al.*, 1993). Conditions were 96°C for 20 sec, 65°C for 40 sec and 70°C for 1 min, for 32 cycles. This resulted in the amplification of the expected 123 bp product on chromosomes 7, 9, 11, 16 and 17, shown in figure 3.7. Some of these amplicons could have arisen from internal loci rather than adjacent to a telomere. To determine this, PCR amplification was carried out with CB0001-24cA and the TelC telomere primer on the monochromosome hybrid panel, followed by Southern hybridisation with probe CB000124c. A characteristic smear of hybridising products with an underlying banding pattern indicated that this sequence was present next to an array of telomere repeats on chromosome 11, shown in figure 3.8. Amplification of products of up to 4 kb strongly suggest that this is a real telomere rather than an interstitial telomere-like repeat array. There was also weaker amplification of a smear of products with an underlying pattern of discrete bands from chromosomes 7 and 17. The discrete bands suggest the presence of highly degenerate repeat arrays, which is a feature of interstitial telomere-like repeats (Wells *et al.*, 1990). However, products of at least 2 kb were detected indicating that a terminal location on these chromosomes is more likely. Amplification of discrete bands of less than 1.3 kb on chromosomes 9 and 16 are likely to have arisen from copies of this subterminal repeat adjacent to interstitial telomere-like repeats. However, the CB0001-24c subterminal repeat may not be present at these chromosome ends (7, 11 and 17) in all individuals. It is highly likely, given the information derived from other subterminal repeat families, that the distribution of this sequence at chromosome ends is polymorphic for copy number and chromosome location, and it is likely to vary between individuals and populations.

3.5 The PGSEGB subterminal repeat

Group 2 clones all had 28 bp of identical telomere-adjacent sequence, which was identified as the telomere-junction region of the PGSEGB subterminal repeat. This repeat resides on a number of normal chromosome ends (Cross *et al.*, 1990), and is likely to be present in a large number of clones in most subtelomeric libraries due to efficient amplification of the 28 bp sequence adjacent to a telomere. At least ten different repeat interspersion patterns were identified, suggesting that these

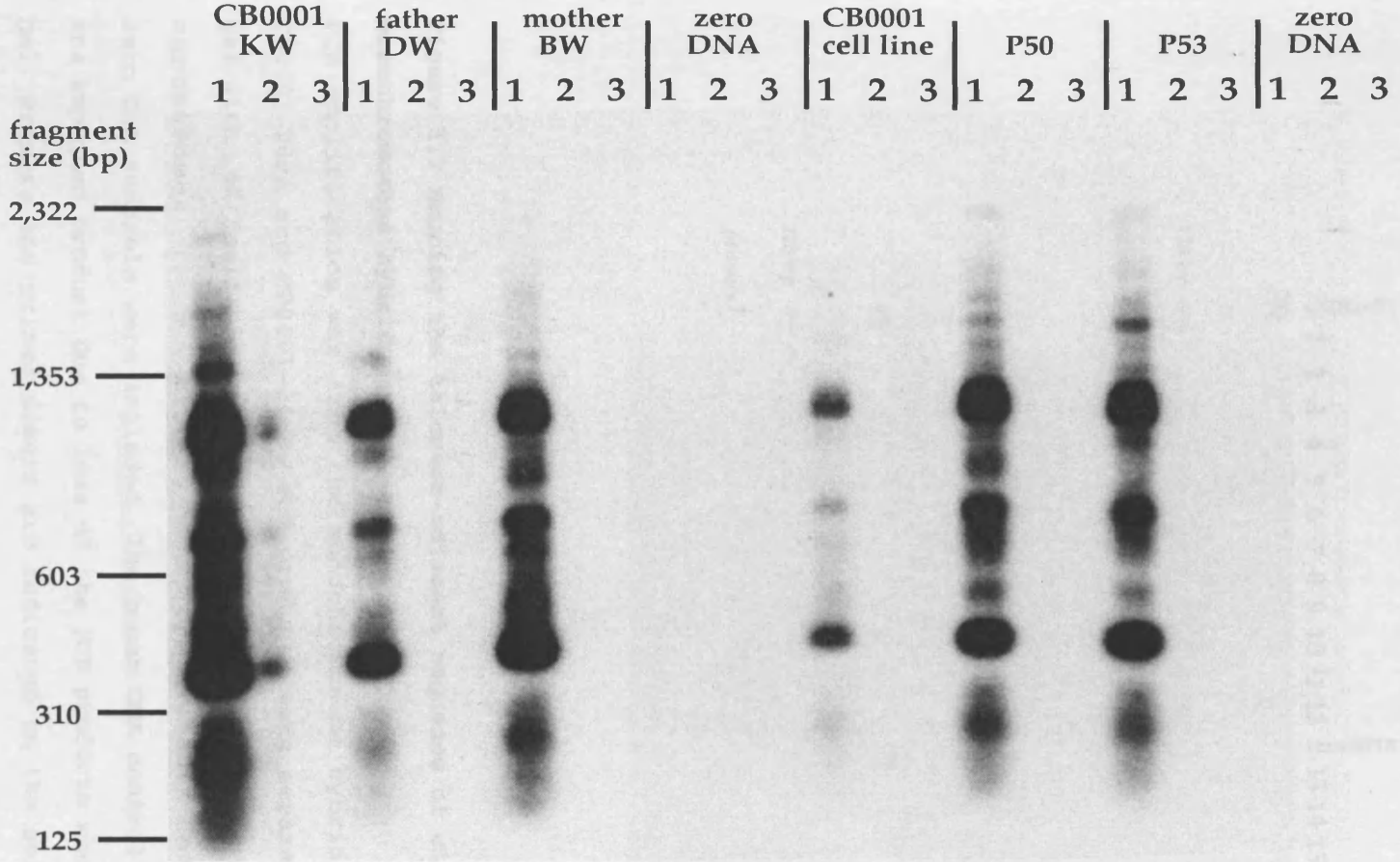


Figure 3.6 Amplification of subterminal repeat CB0001-24c adjacent to a telomere in unrelated individuals

PCR amplification and Southern hybridisation were carried out using primers CB0001-24cA and TelC, and probe CB0001-24c. Detection of characteristic telomere smears in all individuals tested (other individuals not shown) indicates that this sequence is located next to telomere repeats in most individuals. Individuals tested in each reaction are indicated above each triplet of lanes. Lanes 1-3 indicate PCR reactions carried out with 1) both CB0001-24cA and TelC primers, 2) primer CB0001-24cA only and 3) primer TelC only. Reactions 2 and 3 were carried out for control purposes. Products detected in lane 2 of CB0001 are due to leaking of the adjacent well.

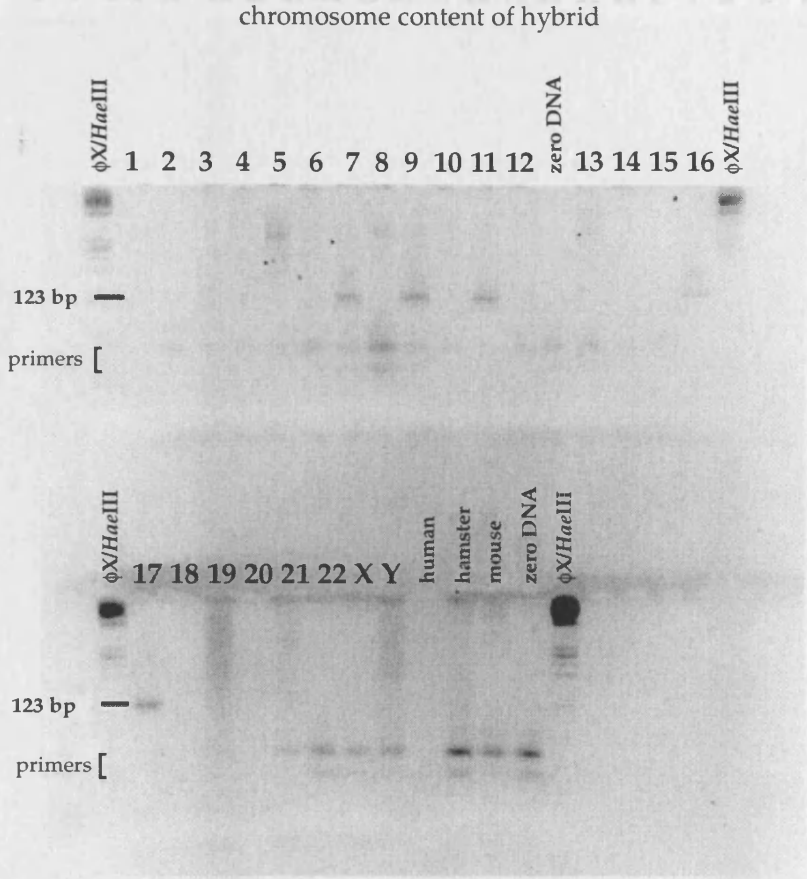


Figure 3.7 Mapping the telomere-adjacent sequence of clone CB0001-24c on the monochromosome hybrid panel

PCR amplification was from the monochromosome hybrid panel using primers CB0001-24cA and CB0001-24cB. PCR products were separated on a 4 % Nusieve gel with φX/*Hae*III DNA markers. The expected 123 bp product amplifies from chromosomes 7, 9, 11, 16 and 17. Human, hamster, mouse and zero DNA controls were included. The human DNA control lane did not contain the expected product due to loss of the PCR products when loading the agarose gel. Primers and primer dimers are indicated on the gel.

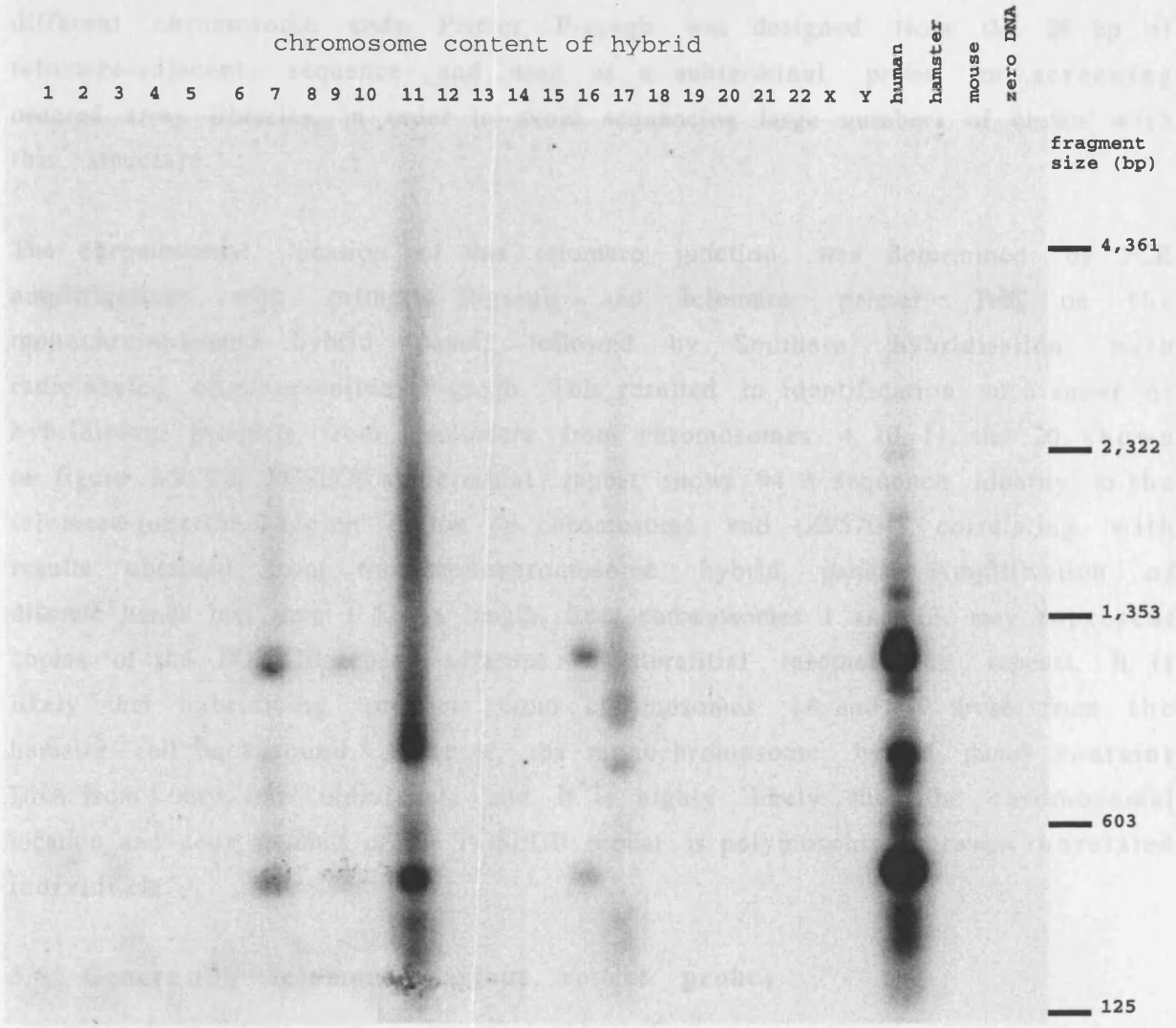


Figure 3.8 Mapping the location of the CB0001-24c telomere on the monochromosome hybrid panel

PCR amplification was from the monochromosome hybrid panel using primers CB0001-24cA and TelC. Human, hamster, mouse and zero DNA control reactions were included. Products were detected by Southern hybridisation using probe CB0001-24c. A smear of hybridising products with an underlying banding pattern was detected from chromosomes 7, 11 and 17. Amplification of discrete bands less than 1.3 kb from chromosomes 9 and 16 probably represents copies of this sequence adjacent to an interstitial telomere-like repeat array.

clones represent at least ten different isolates of the PGSEGB sequence family adjacent to a telomere repeat array, and may be different alleles or isolates from different chromosome ends. Primer P-gsegb was designed from the 28 bp of telomere-adjacent sequence and used as a subterminal probe for screening ordered array libraries, in order to avoid sequencing large numbers of clones with this structure.

The chromosomal location of this telomere junction was determined by PCR amplification with primer P-gsegb and telomere primer TelC on the monochromosome hybrid panel, followed by Southern hybridisation with radiolabeled oligonucleotide P-gsegb. This resulted in identification of a smear of hybridising products from a telomere from chromosomes 4, 10, 11, and 20, shown in figure 3.9. The PGSEGB subterminal repeat shows 94 % sequence identity to the telomere-junction region of the 4p chromosome end (Z95704), correlating with results obtained from the monochromosome hybrid panel. Amplification of discrete bands less than 1 kb in length, from chromosomes 1 and 13, may represent copies of the PGSEGB repeat adjacent to interstitial telomere-like repeats. It is likely that hybridising products from chromosomes 14 and 19 arose from the hamster cell background. However, the monochromosome hybrid panel contains DNA from only one individual, and it is highly likely that the chromosomal location and copy number of the PGSEGB repeat is polymorphic between unrelated individuals.

3.6 Generating telomere variant repeat probes

Sequence data obtained from CB0001 clones established that the majority of telomere positive clones contained variant repeat types TCAGGG, TGAGGG and TTGGGG within the repeat array. The accumulation of variant repeats at the proximal ends of telomeres probably occurs over time, via mutation processes such as replication slippage and intra-allelic exchange events (Baird *et al.*, 1995). Consequently telomeres containing variant repeats can be eliminated as *de novo* repeat arrays. Exclusion of clones containing variant repeats from ordered array libraries prior to sequencing would reduce the number of telomere-junction clones to sequence. Telomere variant repeat probes TelG (TGAGGG), TelJ (TTGGGG) and TelK (TCAGGG) were generated as described in chapter 2 (Materials & Methods - Generating telomere repeat probes). These probes were radioactively labeled and hybridised to replicate filters of CB0001 library plates 1 and 2, and positives scored for each repeat probe.

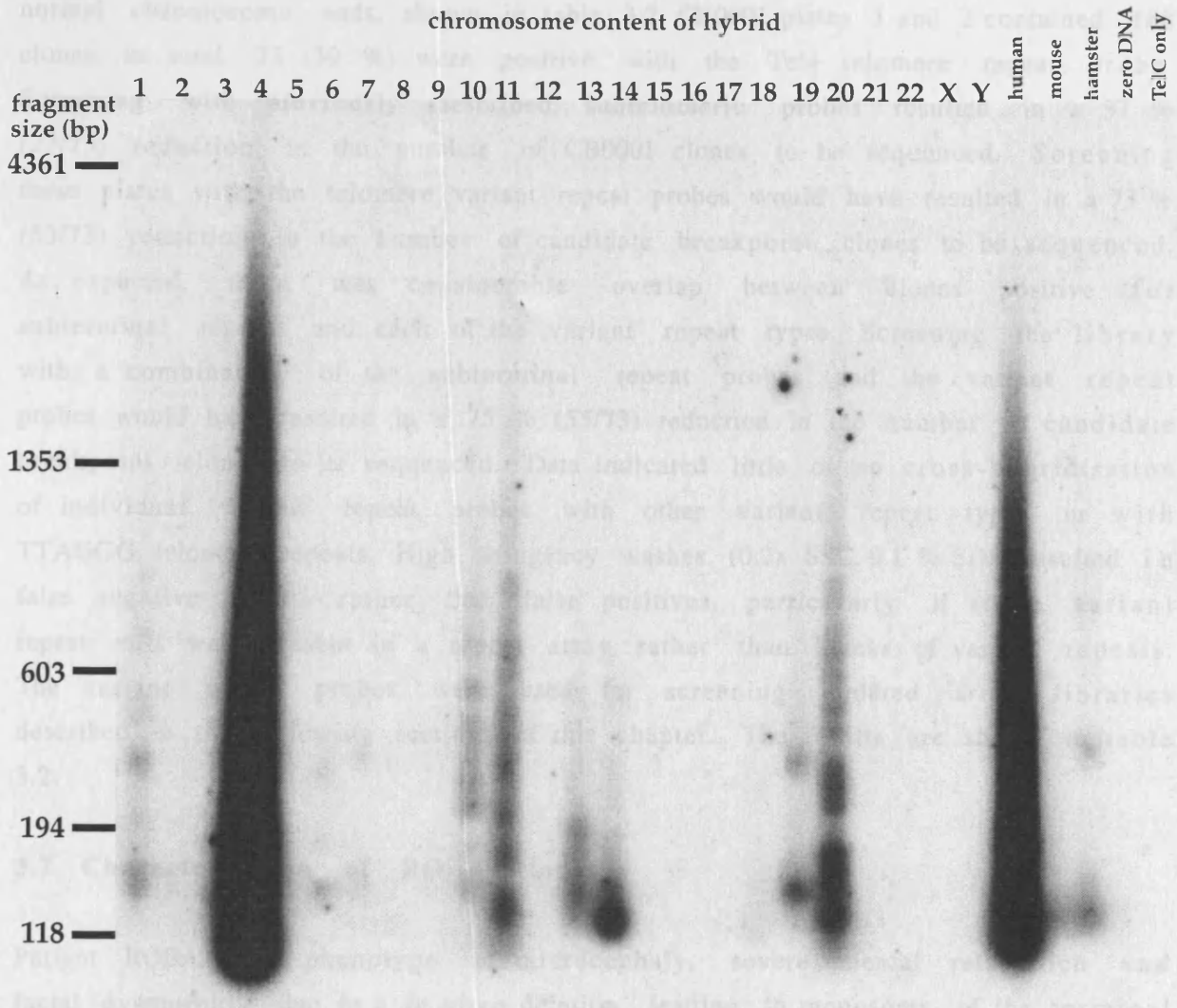


Figure 3.9 Mapping the PGSEGB subterminal repeat on the monochromosome hybrid panel

PCR amplification was from the monochromosome hybrid panel using primer CB0001-28 and telomere primer TelC, and products were detected by Southern hybridisation using oligonucleotide CB0001-28. Human, hamster, mouse and zero DNA controls were included. Detection of an extensive smear of hybridising products characteristic of telomere amplification indicates that the PGSEGB subterminal repeat is found next to a telomere on chromosomes 4, 10, 11 and 20. Chromosomes 1 and 13 show amplification of discrete bands, which may represent copies of this repeat adjacent to interstitial telomere-like repeats. Weak amplification from chromosomes 6, 7, 14 and 19 may be due to hamster DNA background.

Comparison of hybridisation data with sequence data previously obtained from the CB0001 library, indicated that these probes were highly effective in identifying clones containing variant repeat types, including clones arising from many normal chromosome ends, shown in table 3.2. CB0001 plates 1 and 2 contained 188 clones in total. 73 (39 %) were positive with the Tel+ telomere repeat probe. Screening with previously described subtelomeric probes resulted in a 37 % (27/73) reduction in the number of CB0001 clones to be sequenced. Screening these plates with the telomere variant repeat probes would have resulted in a 73 % (53/73) reduction in the number of candidate breakpoint clones to be sequenced. As expected, there was considerable overlap between clones positive for subterminal repeats and each of the variant repeat types. Screening the library with a combination of the subterminal repeat probes and the variant repeat probes would have resulted in a 75 % (55/73) reduction in the number of candidate breakpoint clones to be sequenced. Data indicated little or no cross-hybridisation of individual variant repeat probes with other variant repeat types, or with TTAGGG telomere repeats. High stringency washes (0.2x SSC, 0.1 % SDS) resulted in false negative results rather than false positives, particularly if single variant repeat units were present in a repeat array rather than blocks of variant repeats. The variant repeat probes were used for screening ordered array libraries described in the following sections of this chapter. The results are shown in table 3.2.

3.7 Characterisation of ROBA clones

Patient ROBA has a phenotype of microcephaly, severe mental retardation and facial dysmorphia due to a *de novo* deletion, leading to monosomy of the terminal 5-7 Mb of 1p (Giraudeau *et al.*, 1997). The ROBA telomere-anchored PCR library contained a total number of 658 clones. The Tel+ telomere repeat probe detected 77 (12 %) positive clones (see table 3.2). After screening with subtelomeric probes Tsk37, Tsk46, Tsk8, E-F, CB0001-24c and PGSEGB, and variant repeat probes TelG, TelJ and TelK (see table 3.2), six (8 %) candidate breakpoint clones remained to be sequenced. All clones contained the predicted sequence from the *Kpn*I to *Sau*3A sites of the Sau3AI-linker unless otherwise stated. A summary of clones sequenced from the ROBA library is shown in table 3.3.

Group 1. Four clones contained *Kpn*I sites and telomere repeats only.
R-112g had an array of TTAGGG and degenerate repeats but not variants TCAGGG, TGAGGG or TTGGGG.
R-79b had 289bp of telomere-adjacent sequence but did not show identity to sequences in the EMBL/GenBank database, or to previously isolated chromosome

LIBRARY	Total library size	No. Tel+	% Tel +	No. positive with subterminal probes	No. positive for TelG	No. positive for Tel J	No. positive for TelK	No. positive for any of G/J/K	Total reduction of clones to sequence	No. clones remaining to be sequenced
CB0001	188	73	39	27 (37 %)	25 (34 %)	28 (38 %)	12 (16 %)	50 (69 %)	50 (73 %)	18
ROBA	658	77	12	50 (65 %)	14 (18 %)	28 (36 %)	16 (20 %)	46 (60 %)	71 (92 %)	6
CB	376	129	34	84 (65 %)	44 (34 %)	72 (56 %)	36 (28 %)	101 (78 %)	121 (94 %)	8
CB00054	752	204	27	104 (51%)	86 (42 %)	112 (55 %)	41 (20 %)	167 (82 %)	199 (98 %)	5

Table 3.2 Screening CB0001 and other libraries with variant repeat probes

CB0001 library plates 1 and 2 were screened with variant repeat probes TelG (TGAGGG), TelJ (TTGGGG) and TelK (TCAGGG). Data obtained from this screening step was then compared with data obtained from sequence analysis of these clones. Library name and the total number of clones in each library is shown in the first two columns. All libraries were screened with subterminal probes Tsk37, Tsk46, Tsk8, E-F, CB0001-24c (except CB0001) and PGSEGB (except CB0001) and results are shown in column five. The number of clones containing telomeres that were also positive for each of the variant repeat probes, and the number of clones containing one or more of the variant repeats is shown. The final two columns show the reduction in the number of clones to be sequenced when subterminal and variant repeat probes are used for screening libraries, and the number of remaining left to sequence in each library. Percentages are shown to the nearest whole figure.

Clone name	No. of clones in group	Start site	Telomere-adjacent sequence	Sequence identity FASTA/BLAST database search	Telomere repeats present	Variant repeats present	No. of repeat interspersion patterns in group	Previously identified?
GROUP 1	4	*Kpn* I	no	_	yes	no	_	_
112g	1	*Sau* 3A	no	_	yes	variant repeats	_	_
79b	1	*Sau* 3A	289 bp	no significant matches	yes (12)	TTGGGG other variants	_	no
75h	1	*Sau* 3A	> 200 bp	93% 9q subterminal region (Z96777) 91 % Interleukin 9 receptor pseudogenes (L39064)	yes	TCAGGG	_	no

Table 3.3 Summary of clones sequenced from the ROBA library

Clone names are shown in the first column. Clones are grouped according to similarity either in the telomere-adjacent DNA or in the telomere repeat array, if adjacent DNA is absent. The number of clones in each group, and the start site, *Sau* 3A (GATC) or *Kpn* I (GGTACC), are shown in columns two and three. The length of telomere-adjacent sequence in each group of clones is shown in basepairs, and results of sequence homology searches using BLAST and FASTA search tools at the EMBL/Genbank database are shown in column five. Presence or absence of telomere repeats, TGAGGG, TTGGGG, TCAGGG and other variant repeats is shown in columns six and seven. The number of interspersion patterns within the telomere repeat arrays of each group of clones is indicated in column eight. The final column indicates whether a clone has been previously isolated and from which ordered array library.

ends. This clone had an array of twelve repeats, including one TTGGGG repeat, and other variant repeats, and was thought to represent a novel telomere-junction from a normal chromosome end.

A number of clones positive for variant repeat probes were also selected for sequencing in order to look at repeat interspersion patterns. Most of these contained telomere repeat arrays and variant repeats only and were not investigated further. **Clone R-75h** had telomere-adjacent sequence with 93 % sequence identity to a region unique to the 9q telomere (EMBL/GenBank accession no. Z96777; NIH/IMM, 1996) and 91 % sequence identity to the Interleukin 9 receptor pseudogenes (L39064) located in the subterminal regions of 9qter, 10qter, 16pter and 18pter (Kermouni *et al.*, 1995). This clone had a telomere repeat array containing TCAGGG variant repeat types.

No candidate terminal-deletion breakpoint clones were identified in the ROBA library.

3.8 Further investigation of ROBA clone R-79b

Clone R-79b had 289bp of telomere-adjacent sequence without identity to sequences in the EMBL/GenBank database, or previously isolated chromosome ends. The telomere repeat array containing nine 6 bp repeats as follows: (ttaggg)(ttgggg)(ttaggg)(ctgggg)(ttaggg)(ttagag)(tacgg)$(ttaggg)_2$. Primers R79bA and R79bB were designed from the telomere-adjacent sequence to generate a 168 bp product. This region was mapped on the monochromosome hybrid panel, and the expected 168 bp product mapped to chromosome 21 only, shown in figure 3.10. PCR amplification on a panel of patient and unrelated DNAs (P50, P53, P55, CEPH parents 133401, 133402, 10202 and 0201), with flanking primer R-79bA and the TelC telomere primer, resulted in the amplification of a discrete product of approximately 280 bp in all individuals (see figure 3.11), rather than the smear of hybridising products characteristic of amplification into a telomere. The TelC telomere primer contains three full TTAGGG repeats and would therefore be expected to be efficient at priming from only the terminal two repeats in the cloned telomere repeat array in R-79b, due to the presence of variant repeat types. The R-79bA telomere-adjacent primer begins 194 bp from the start of the telomere repeat array, and in conjunction with TelC would be expected to generate products ranging from 279 to 285 bp, including the TelC non-complementary tail. This expected product size correlates well with the 280 bp product amplified from genomic DNA using these primers. Given these results, it is likely that clone R-79b contains the entire array of telomere repeats found at this locus, and represents a

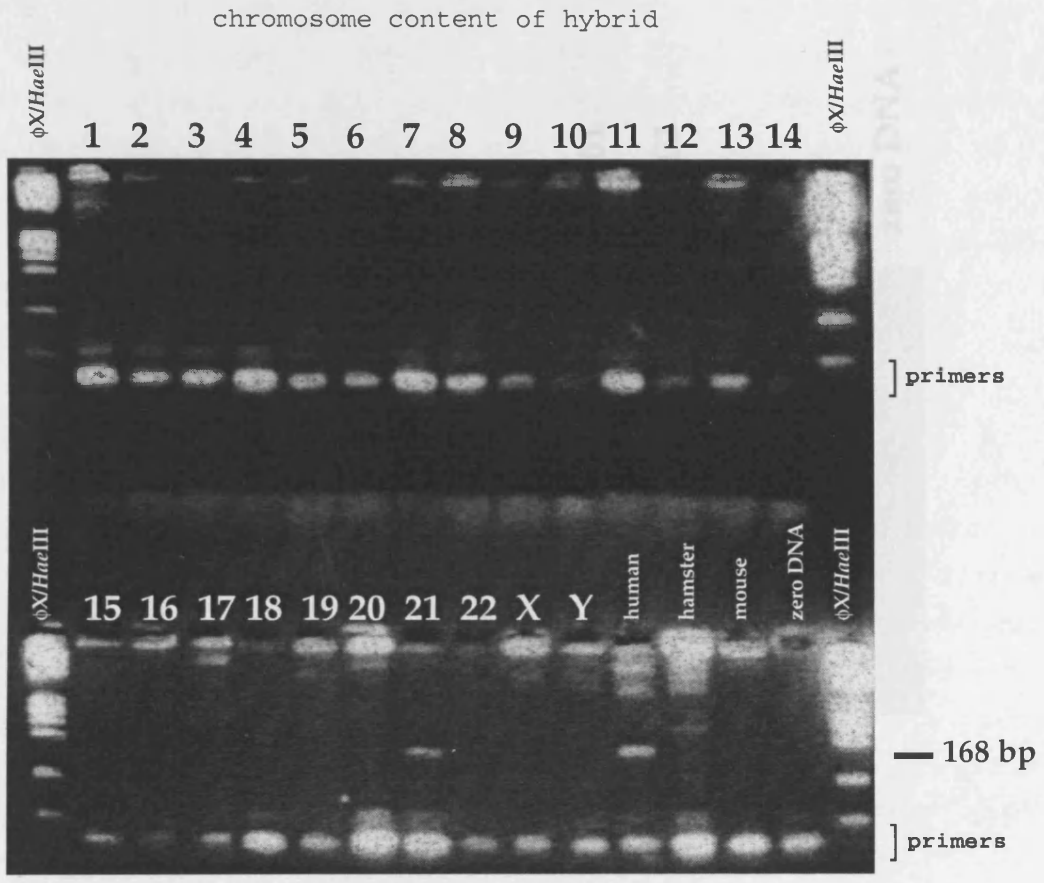


Figure 3.10 Mapping the telomere-adjacent sequence of clone R-79b on the monochromosome hybrid panel

PCR was from the monochromosome hybrid panel using primers R79bA and R79bB. Human, hamster, mouse and zero control DNAs were included. PCR products were separated on a 4 % Nusieve agarose gel with φX/*Hae*III DNA markers. The correct 168 bp product amplifies from chromosome 21 only. The small bands in each lane represent primers and primer dimers.

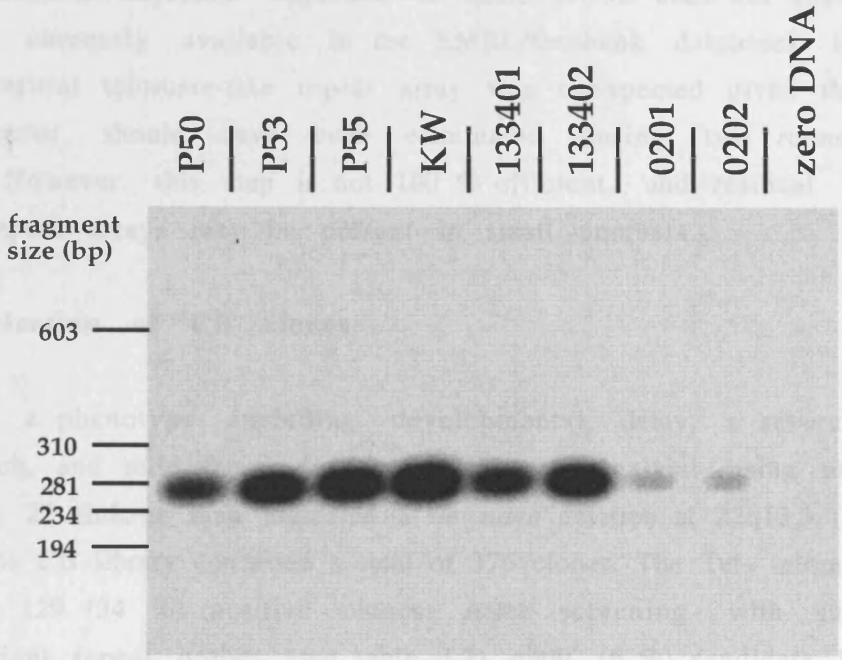


Figure 3.11 Amplification of the telomere-adjacent DNA and telomere-like repeats found in clone R-79b in unrelated individuals

PCR amplification was carried out using using primers R-79bA and TelC, and products detected by Southern hybridisation with probe R-79b. Individuals used are indicated above each lane. Amplification of a discrete 281 bp product in all individuals tested, including patient ROBA (data not shown), indicates that clone R-79b represents a short interstitial array of telomere-like repeats found in most or all individuals.

short internal array of TTAGGG and degenerate repeats, located on chromosome 21, in most or all individuals. Short, degenerate repeat arrays are a feature of interstitial telomere-like repeats (Wells *et al.*, 1990). These interstitial telomere-like repeats are known to share sequence similarity outside the repeat array, although the telomere-adjacent sequence in clone R-79b does not share identity with sequences currently available in the EMBL/Genbank databases. Isolation of this short interstitial telomere-like repeat array was unexpected given that the 280 bp *Mbo*I fragment should have been eliminated during two rounds of size fractionation. However, this step is not 100 % efficient, and residual interstitial telomere-like repeat arrays may be present in small numbers.

3.9 Characterisation of CB clones

Patient CB has a phenotype including developmental delay, a severe delay in expressive speech, and mild facial dysmorphia. Dosage analysis using probes from the chromosome 22 linkage map indicated a *de novo* deletion at 22q13.3 (Nesslinger *et al.*, 1994). The CB library contained a total of 376 clones. The Tel+ telomere repeat probe detected 129 (34 %) positive clones. After screening with subtelomeric probes and variant repeat probes (see table 3.2), eight (6 %) candidate breakpoint clones remained to be sequenced. All clones contained the predicted sequence from the *Kpn*I to *Sau*3A sites of the Sau3AI-linker unless otherwise stated. A summary of clones sequenced from the CB library is shown in table 3.4.

Group 1. Four clones had arrays of telomere repeats only, a fifth clone had an array of telomere repeats and variant repeats but not TGAGGG, TCAGGG or TTGGGG.
Group 2. Two identical clones had a *Kpn*I start site and showed 99 % sequence identity with part of the PGB4G7 subterminal repeat (X56278; Cross *et al.*, 1990). The telomere repeat arrays were also identical and contained variant repeats.
CB-112e had a *Kpn*I start site, 13 bp of telomere-adjacent sequence and an array of telomere repeats including TTGGGG repeats. However, this clone was negative for TelJ, probably due to only singleton variant repeats in the array, and consequent inefficient hybridisation to the TelJ probe.
CB-46a had 49 bp of telomere-adjacent DNA which showed 100 % identity to the telomere-adjacent sequence of the CB0001 polymorphic telomere described in section 3.4. This clone had an array of telomere repeats at the same location as the polymorphic telomere, and contained two variant repeats (TTAAGGG). The interspersion pattern of repeats in this telomere is different to that seen in the CB0001 polymorphic telomere, as expected due to the extensive variation seen at other telomeres.

Clone name	No. of clones in group	Start site	Telomere-adjacent sequence	FASTA/BLAST database search results	Telomere repeats present	Variant repeats present	No. of repeat interspersion patterns in group	Previously identified?
GROUP 1	6	*Sau* 3A	none	–	yes	variant repeats in 1 clone	–	–
GROUP 2	2	*Kpn* I	> 200 bp	99 % PGB4G7 subterminal repeat (X56278)	yes	variant repeats	1	NT
GROUP 3	5	*Sau* 3A	47 bp	92 % subterminal tandem repeat (U53226)	yes	TCAGGG, TGAGGG other variants	2	no
GROUP 4	2	*Sau* 3A	47 bp	100 % L1 repeat (AC005248)	yes	TCAGGG TGAGGG other variants	1	no
112e		*Kpn* I	13 bp	–	yes	TTGGGG other variants	–	no
46a		*Sau* 3A	49 bp	96 % 22q subterminal repeat (AC002055) 94 % 10q/4q subterminal repeat (AF017467) 90 % 4ptel (Z95704)	yes	variant repeats	–	CB0001 polymorphic telomere
12g		*Kpn* I	> 200 bp	100 % L1 repeat at Xq (AL022151) and chromosome 5 (AC004769)	no	–	–	no

Table 3.4 Summary of clones sequenced from the CB library

Clone names are shown in the first column. Clones are grouped according to similarity either in the telomere-adjacent DNA, or telomere repeat array if adjacent DNA is absent. The number of clones in each group, and the start site, *Sau* 3A (GATC) or *Kpn* I (GGTACC), are shown in columns two and three. The length of the telomere-adjacent sequence in each group of clones is shown in basepairs. Results of sequence homology searches using BLAST and FASTA search tools at the EMBL/Genbank database are shown in the fifth column. Presence or absence of telomere repeats, and TCAGGG, TTGGGG, TGAGGG or other variant repeats is shown in columns six and seven. The number of interspersion patterns within the telomere repeat arrays of each group of clones is indicated in column eight. The far right-hand column indicates whether a clone has been previously isolated and from which ordered array library.

A number of clones positive for variant repeat probes were also selected for sequencing to investigate repeat interspersion patterns. **Group 3.** Five clones had 46bp of identical telomere-adjacent sequence with 92 % sequence identity to a subterminal tandem repeat (U53226), found at 3q, 4p and on acrocentric chromosomes 13, 14, 15, 21 and 22 (Thoraval *et al.*, 1996). These clones had arrays of telomere repeats containing variants including TCAGGG, TGAGGG and other variant repeats. The interspersion patterns of repeats in the telomere repeat array were identical in four clones and different in a fifth clone, representing four independent isolates from one location, and a single isolate of the sequence from another allele or different chromosomal location. **Group 4.** Two clones had 47 bp of identical telomere-adjacent sequence with 100 % sequence identity to an L1 repeat, and identical arrays of telomere repeats containing variants TCAGGG and TGAGGG. **CB-12g** had a *Kpn*I start site, and telomere-adjacent sequence showing 100% sequence identity over 144 bp to an L1 repeat found at Xq22.1-22.3 (A1022151) and on chromosome 5 (AC004769). No repeats were detected on sequencing, and this clone probably represents a false positive for the telomere repeat probe.

No candidate terminal-deletion breakpoint clones were isolated from the CB library.

3.10 Characterisation of CB00054 clones

Patient CB00054 has a phenotype including short stature, brachymetaphalangia (shortening of the bones in the hands and feet) and mental retardation (Wilson *et al.*, 1995) due to a cytogenetically visible deletion at 2q37.2-q37.3. The library contained a total of 752 clones. The Tel+ telomere repeat probe detected 204 (27 %) positive clones. After screening with subtelomeric probes and variant repeat probes (see table 3.2), five (2 %) candidate breakpoint clones remained to be sequenced. All clones contained the predicted sequence from the *Kpn*I to *Sau*3A sites of the Sau3AI-linker unless otherwise stated. Clones sequenced from the CB00054 library are summarised in table 3.5.

Group 1. Three clones contained only telomere repeat arrays and were not investigated further.

CB54-38e had more than 200 bp of telomere-adjacent sequence with 98 % sequence identity to an *Alu* element of the Sg-subfamily. This clone was strongly positive with the Tel+ probe but repeats were not evident on the sequencing gel. Positive hybridisation to the telomere repeat probe may have been due to the presence of a second colony in the well containing telomere repeats, which was

Clone name	No. of clones in group	Start site	Telomere-adjacent sequence	FASTA/BLAST database search results	Telomere repeats present	Variant repeats present	No. of repeat interspersion patterns in group	Previously identified?
GROUP 1	3	GATC	none		yes	no		_
38e	1	GATC	> 200 bp	98 % Alu-Sg repeat (Z93023)	no	_		no
21d	1	GATC	265 bp	68 % Alu-Sx repeat	yes	variant repeats		no
38a	1	GATC	212 bp	93 % 7q subterminal repeat (AF027390)	yes	TGAGGG variant repeats		no

Table 3.5 Summary of clones sequenced from the CB00054 library

Clone names are shown in the far left-hand column. Clones are grouped according to similarity either in the telomere-adjacent DNA, or telomere repeat array if adjacent DNA is absent. The number of clones in each group, and the start site, *Sau* 3A (GATC) or *Kpn* I (GGTACC), are shown in columns two and three. The length of telomere-adjacent sequence in each group of clones is shown in basepairs. Results of sequence homology searches using BLAST and FASTA search tools at the EMBL/Genbank database are shown in the fifth column. Presence or absence of telomere repeats, and TGAGGG, TTGGGG, TCAGGG and other variant repeats is shown in columns six and seven. The number of interspersion patterns within the telomere repeat arrays of each group of clones is indicated in column eight. The far right-hand column indicates whether a clone has been previously isolated and from which ordered array library.

not propagated when the contents of this well were grown for single-stranded DNA extraction.

CB54-21d had 265 bp of telomere-adjacent sequence with limited (68 %) sequence identity to an *Alu* element of the Sg and Sx subfamilies. The telomere repeat array contained many variant repeats such as TTACGGG and TTCAGGG, consequently this clone was not investigated further.

A number of clones positive for variant repeat probes were also selected for sequencing. Most contained telomere repeats with variants only. **CB54-38a** had 212bp of telomere-adjacent sequence which showed 93 % sequence identity to a region of the 7q telomere (AF027390), normally located approximately 26 kb from the telomere repeat array. This clone also contained an array of telomere repeats, including TGAGGG variants. CB54-38a may represent a copy of the 7q-like region located on another chromosome end. However, it is known that the low copy subterminal repeats at the 7q telomere extend for only a few kilobases proximal to the telomere. 7q unique sequence has been detected approximately 5.5 to 6.7 kb from the telomere (Brown *et al.*, 1990). This clone may represent a length polymorphism, or a truncation and healing event in the subterminal region of 7q, without a phenotype.

No terminal-deletion breakpoint candidate clones were isolated from the CB00054 library.

Discussion

3.11 Limitations of the telomere-anchored PCR strategy

Cloning of human chromosome ends has enabled characterisation of subterminal repeat regions, and their relationship with other repeat families present at chromosome ends or interstitial locations. Subterminal repeat families are present on different subsets of chromosomes, and often have polymorphic distribution between unrelated individuals (Brown *et al.*, 1990; Cross *et al.*, 1990; de Lange *et al.*, 1990). The telomere-anchored PCR method has been used to isolate a number of telomere-junction sequences, and a terminal deletion breakpoint from patient FB336R. The breakpoint appears to have been healed in the germline by *de novo* addition of telomere repeats. This patient has only a very mild phenotype in comparison to other HPE patients, despite having a large terminal deletion. The mild phenotype seen FB336R may be due to mosaicism, a consequence of chromosome healing very early on during development rather than in the germline. *De novo* telomere addition is likely to be mediated by telomerase, which is active in the germline and at the blastocycst stage. Telomerase remains active until at least 16-20 weeks during fetal development except in brain tissue (Wright *et al.*, 1996b). Therefore healing of a truncated chromosome by telomerase could take place in the germline or early in development.

However, the telomere-anchored PCR strategy was unsuccessful in the isolation of terminal deletion breakpoints from ordered array libraries generated from patients NT, FB241 (N.J. Royle, personal communication), CB0001, CB, CB00054 and ROBA. There were a number of potential problems with the strategy. Firstly, the availability of patients with known terminal deletions is limited. The CB deletion was identified by high resolution cytogenetic analysis and the extent of the deletion mapped by dosage analysis using probes from the most complete linkage map of chromosome 22 available at the time. The most distal marker, ARSA, was deleted, however, it was unknown how far this marker was from the telomere (Nesslinger *et al.*, 1994). The CB00054 (2q) microdeletion was not cytogenetically visible and was mapped by microsatellite marker analysis. The CB0001 deletion on 2q was cytogenetically visible, but the proximal extent of the deletion could not be defined precisely, and lay within a 22 cM region. The most telomeric informative marker on 2q at this time was D2S125, but it was not known how far from the telomere this marker was located (Wilson *et al.*, 1995). The ROBA deletion was determined by reverse chromosome painting, microsatellite marker analysis and Southern blot analysis. The deletion extended from the most telomeric marker CEB108, which is 0.2 cM from the telomere (Giraudeau *et al.*, 1997). The possibility

that these patients have interstitial deletions or a translocation involving subtelomeric repetitive DNA cannot be ruled out. If the deletions in these patients are interstitial, it is likely that the existing telomere remains at the chromosome end, with some subterminal sequence remaining adjacent to the telomere. Similarly, broken chromosomes stabilised by capture of an existing telomere, or by copying of a telomere from another chromosome, would be likely to have subterminal sequences remaining adjacent to the telomere, and variant repeats within the telomere repeat array. The screening criteria used to select candidate breakpoints from ordered array libraries would not detect this type of healing event, as it is dependent on the assumption that the *de novo* telomere addition results in complete loss of the subterminal region. It is possible that even very large cytogenetically visible deletions, such as FB241 and CB00054 are interstitial, with subterminal DNA remaining adjacent to the telomere. Due to the polymorphic nature of subterminal regions, it has been extremely difficult to generate unique probes for different chromosome ends, which would allow the rapid assessment of the nature and extent of suspected terminal deletions. A FISH technique has been developed involving 41 telomere-specific probes, allowing simultaneous analysis of the subtelomeric regions of every chromosome for deletion, triplication and balanced translocation events (Knight *et al.*, 1997). This will be a very useful diagnostic tool in future for characterising chromosomal rearrangements. However, there is no way to detect terminal deletions with complete certainty. These unique probes are located 100-300 kb from the end of the chromosome (NIH/IMM, 1996), and would not identify deletions where a small region of subterminal sequence remains adjacent to the telomere.

Other potential problems with the strategy include the efficiency of the telomere-anchored PCR step. Amplification may be suboptimal, that is, occurring from a only a small number of chromosome ends rather than a representative proportion of *Mbo*I digestion products, potentially excluding a breakpoint fragment. If a breakpoint is located within a region that might be refractory to PCR amplification, such as a minisatellite or other repeated region, PCR amplification of the region may also be inefficient. Similarly, if the *Mbo*I restriction fragments adjacent to the breakpoint are long, they are unlikely to be amplified to high copy number in the telomere-anchored PCR reaction. Cloning of large or low copy number PCR products is very inefficient, due to the preferential cloning of smaller, high copy number fragments generated during the PCR step, and again may lower the probability of cloning a breakpoint fragment. Analysis of clone insert size in previously characterised libraries indicated that the majority of inserts were less than 1.5 kb. The breakpoint in patient NT was not isolated from an ordered array library generated using the telomere-anchored PCR method (N.J.

Royle, unpublished data). This breakpoint has been cloned through contig analysis and found to lie within minisatellite locus MS607 (Wong *et al.*, 1997). The ordered array telomere-anchored PCR library was re-screened with a probe to detect the NT breakpoint (N.J. Royle personal communication), but no positive clones were identified. However, the terminal *Mbo*I breakpoint fragment (3.5 kb) was present at very low levels in the telomere-anchored PCR products. A number of factors may have contributed to failure of cloning this breakpoint, including the low level of PCR amplification, the large size of the terminal deletion breakpoint fragment (3.5 kb) and observed instability of cloned minisatellite repeat arrays (N. J. Royle, personal communication).

3.12 Modifying the library screening step

The work described in this chapter has resulted in a number of improvements to the library screening step, and the generation of telomere-adjacent sequence data. Two common subterminal repeat families were isolated from library CB0001, the PGSEGB subterminal repeat (Cross *et al.*, 1990), and a novel subterminal repeat family represented by clone CB0001-24c. At least ten independent isolates of the PGSEGB repeat family adjacent to a telomere repeat array were isolated, representing different alleles or copies of this repeat from different chromosome ends. Mapping the PGSEGB sequence on the monochromosome hybrid panel identified telomere-adjacent copies on chromosomes 4, 10, 11, and 20. The CB0001-24c subterminal repeat was found to be present adjacent to an array of telomere repeats in all 21 individuals tested. Copies of this repeat were located adjacent to a telomere on chromosomes 7, 11 and 17. However, it is likely that the chromosomal location and copy number of the PGSEGB and CB0001-24c subterminal repeats are polymorphic between unrelated individuals and have different frequencies among different populations. Sequence data from these subterminal repeat families was used to generate probes PGSEGB and CB0001-24c for screening ordered array libraries to eliminate these normal chromosome ends prior to sequencing.

Variant repeat screening was developed to reduce unnecessary sequencing of normal chromosome ends containing variant repeats TGAGGG, TCAGGG and TTGGGG. The variant repeat probes were highly effective in identifying telomere repeat arrays containing blocks of these variant repeat types, and screening with these probes in conjunction with the subterminal repeat probes resulted in a significant reduction in the number of candidate breakpoint clones to investigate, making library analysis much less time consuming. This data is summarised in table 3.1.

A database of subterminal sequence was compiled using sequence data generated from analysis of ordered array libraries described in this chapter. All novel chromosome ends were entered into the database. Information includes clone start site (*Sau*3A or *Kpn*I), complete telomere-adjacent sequence, and indication of the start of the telomere repeat array. The database can be queried using a FASTA search program on the Genetics Computer Group (GCG) sequence analysis software (IRIX release 6.4), and updated with additional sequence when further chromosome ends or subterminal regions are isolated. New sequence is queried against this database to prevent unnecessary investigation of normal chromosome ends that have been previously isolated.

3.13 Chromosome length polymorphism

Chromosome length polymorphism has been identified at the tip of 16p, 16q (Wilkie *et al.*, 1991) and 2q (Macina *et al.*, 1994), where the telomere is located at different distances from a fixed locus on different alleles. A chromosome end with a telomere that is polymorphic for presence or absence in unrelated individuals was identified in patient NT (Coleman *et al.*, 1999). This telomere was unusual due to the truncation of a subterminal region by telomere repeats at a more internal location than in most individuals. This polymorphic telomere is low frequency, ranging from 4.5 % to 8.6 % in different populations, and is inherited in a Mendelian fashion. The telomere-adjacent DNA has high sequence identity to ESTs usually found in subterminal regions but not adjacent to telomeres at 22q (AC002055), 4q/10q (U74496) and 4p (Z95704). The polymorphic telomere has relocated at least once since it arose, and linkage analysis suggests it is present at 16q at a frequency of 6 % in the Caucasian population. It may also be linked to 16p, but does not show identity with the four alleles at 16p previously identified (Wilkie *et al.*, 1991) and may represent another allele.

The polymorphic telomere identified in library CB0001 is present at very low frequency (2.4 %) in the Caucasian population. The telomere-adjacent DNA also shows high sequence similarity to ESTs found in subterminal regions of 22q (AC002055), 4q/10q (AF017467; U74496) and 4p (Z95704), although this similarity is found in different regions of these subterminal sequences to that of the NT polymorphic telomere. The chromosomal location of the CB0001 polymorphic telomere is unknown but could be determined by linkage analysis, and the terminal location confirmed by *Bal*31 digestion, an exonuclease that degrades the 5′ and 3′ termini of duplex DNA. Presence of the CB0001 polymorphism in the proband's father suggests that it is inherited in a Mendelian manner, but this has not been investigated further. CEPH individual 133101 was also positive for this

polymorphic telomere, and the availability of DNA from family members would allow investigation of the mode of inheritance.

These polymorphisms probably arose via truncation of this subterminal region at different locations by a telomere repeat array. Possible mechanisms include healing of a broken chromosome by telomerase, or an interstitial deletion, moving an existing telomere to a more proximal location. These truncation and healing events may be a common occurrence in subterminal regions. A number of telomere-junction clones were isolated from different libraries with sequence identity to different regions of the same subterminal repeat. For example, group 5 clones from the CB0001 library, and group 2 clones from the CB library show strong sequence identity to different regions of the PGB4G7 subterminal repeat. These clones suggest that variable truncation of the PGB4G7 subterminal repeat by a telomere repeat array has occurred. Similar observations have been made in other subterminal repeat families, such as Tsk37 (reviewed Royle, 1995).

It is likely that chromosome length polymorphism without phenotype is present at other chromosome ends. Subterminal regions do not appear have a function with respect to chromosome stability as chromosomes with terminal deletions causing complete loss of the subterminal region, healed by *de novo* telomere addition, are stable (Flint *et al.*, 1994; Wilkie *et al.*, 1990). Subterminal sequences may merely act as buffer to prevent loss of coding sequence from the end of the chromosome due to small terminal deletions.

Chapter 4

INVESTIGATION OF THE TELOMERE-ANCHORED PCR AND CLONING STEPS USED TO CONSTRUCT ORDERED ARRAY LIBRARIES

Summary

Efficient cloning of telomere-junction fragments to construct ordered array libraries was not always reproducible, and often low, limiting the success of this strategy. Potential technical problems included inefficient telomere-anchored PCR amplification from some chromosome ends, resulting in the cloning of only a small population of telomeres, which may exclude breakpoint fragments. An ordered array library was generated using pooled genomic DNA from 15 unrelated CEPH individuals. The distribution of chromosome ends in this library was compared with other libraries characterised. Data indicated that telomere-anchored PCR consistently and efficiently amplified from chromosomes Xp/Yp, 4, 7, 10, 11, 12, 17, 18, 20, and other chromosome ends comprised of subterminal repeat families with polymorphic distribution. Inefficient amplification of digested fragments may arise if the terminal *Mbo*I site is more than 1 kb from the telomere, due to the limitations of PCR amplification. Long-PCR was optimised and shown to result in significant enrichment for a 3.5 kb *Mbo*I terminal deletion breakpoint fragment in individual NT, although this product was not cloned successfully, probably due to size or internal structure. Long-PCR was used to generate telomere-junction libraries from other patients, but resulted in efficient amplification of a heterogeneous group of non-telomeric products, probably originating from incompletely digested genomic DNA fragments. These products were cloned at high frequency in preference to telomere repeat arrays, consequently long-PCR was not pursued further. Cloning of PCR products larger than 1 kb is inefficient due to preferential cloning of smaller products. Size fractionation of PCR products prior to cloning resulted in the isolation of longer fragments, but these clones contained longer telomere repeat arrays rather than longer telomere-adjacent DNA sequence, and cloning efficiency was very low (3 %). Consequently, size fractionation prior to cloning was not pursued further. It was concluded that the telomere-anchored PCR and cloning steps were efficient within the technical limits of the strategy.

Introduction

The telomere-anchored PCR method was used to isolated a terminal deletion breakpoint from individual FB336R. This method was applied to isolate breakpoints from six patients (NT, FB241, CB0001, CB, CB00054 and ROBA) with suspected terminal deletions on other chromosomes, but was not successful. A number of potential problèms were identified, and were described in detail in chapter 3. Firstly, it was possible that patients ROBA, CB, CB0001 and CB00054 had interstitial deletions rather than terminal deletions, and therefore the breakpoints would not be identified due to the screening criteria used to identify candidate breakpoint clones. The strategy only allows detection of novel telomere repeat arrays. However, the nature of these terminal deletions cannot be determined with certainty. Secondly, the telomere-anchored PCR amplification was potentially suboptimal, occurring preferentially from only a few chromosome ends rather than from all available chromosome ends. Poor amplification may occur particularly in regions containing complex repeat arrays or GC-rich regions (such as minisatellites or other dispersed repeats) which are clustered in the proterminal regions of chromosomes. This may introduce bias in the population of molecules generated in the PCR reaction, and lead to exclusion of a breakpoint fragment. Similarly, if the most distal *Mbo*I restriction site at the chromosome terminus lies some distance from the telomere, for example more than 1 kb, the breakpoint fragment is unlikely to be amplified to high copy number due to the long fragment size. Poor PCR amplification of a chromosome end leads to low cloning efficiency, due to preferential cloning of smaller, high copy number products generated in the PCR reaction. In order to address these potential technical problems, the telomere-anchored PCR and cloning steps were investigated, and possible improvements tested.

Aims if the work described in this chapter

The aim of the work described in this chapter was to investigate the PCR and cloning steps used to generate ordered array libraries, determine whether they were optimal, and modify them if required. It was anticipated that modifications to these steps would increase the efficiency of the strategy, resulting in a greater probability of isolating terminal-deletion breakpoint fragments in other patients.

Results

Section 1. Telomere-anchored PCR

4.1 Constructing an ordered array library with genomic DNA from 15 CEPH individuals

An ordered array library was generated from genomic DNA from 15 normal CEPH individuals. This library was constructed to isolate additional normal chromosome ends and novel polymorphic telomeres, and to look at the distribution and number of chromosome ends amplified by telomere-anchored PCR when a large number of target chromosome ends were available in the PCR reaction.

4.2 Characterisation of CEPH clones

The library was generated as previously described (Materials and Methods - Construction of Telomere-Anchored PCR libraries) using a total of 30 μg of genomic DNA from 15 randomly selected individuals from the CEPH panel of Caucasian DNAs (133413, 136216, 133112, 10413, 10414, 135013, 135012, 88416, 137712, 137710, 140810, 142110, 14209, 141811 and 141813). Replicate nylon filters were screened with the Tel+ telomere repeat probe and subterminal probes described in chapter 3, and clones of interest sequenced. Variant repeat probes TelG, TelJ and TelK were not used for screening, as they would be likely to exclude novel chromosome ends. The library contained a total of 1504 clones. 969 (64 %) clones were positive with the Tel+ telomere repeat probe. After screening with subterminal repeat probes, 254 (26 %) telomere positive clones remained to be sequenced. A total of 51 clones were sequenced from library plates 1-4 and are summarised in table 4.1. Further clones were not sequenced as most clones analysed contained telomere-repeat arrays only. No novel chromosome ends or potentially polymorphic telomeres were identified. However, sequence data and subterminal probe hybridisation data generated from the CEPH library was compared with data from libraries described in chapter 3.

4.3 Distribution of chromosome ends in libraries characterised to date

It is difficult to calculate the minimum number of independent clones that would have to be picked in order for a library to contain at least one copy of every chromosome end. However, in a library of a given size, and a given percentage of telomere positive clones, it is possible to calculate the probability that all

Clone name	No. of clones identified	Start site	Telomere-adjacent sequence	FASTA/BLAST database search results	Telomere repeats present	Variant repeats present	Sequence identity to previously identified ends
GROUP 1	25	*Sau* 3A	no	–	yes	TTGGGG, TCAGGG	–
GROUP 2	15	*Kpn* I	no	–	yes	TCAGGG other variants	–
GROUP 3	2	*Kpn* I	no	–	yes	no	–
GROUP 4	8	*Sau* 3A	< 22 bp	–	yes	TTGGGG, TCAGGG other variants	–
C-19b		*Sau* 3A	228 bp	96 % PGB4G7 subterminal repeat (X56278	yes	TCAGGG other variants	NT CB0001
C-212f		*Sau* 3A	225 bp	100 % human endogenous retrovirus (U87596)	no	–	no

Table 4.1 Summary of clones sequenced from the CEPH library

Clone names are shown in the first column. Clones are grouped according to similarity either in the telomere-adjacent DNA, or telomere repeat array if adjacent DNA is absent. The number of clones in each group, and the start site, *Sau* 3A (GATC) or *Kpn* I (GGTACC), are shown in columns two and three. The length of telomere-adjacent sequence in each group of clones is shown in basepairs. Results of sequence homology searches using BLAST and FASTA search tools at the EMBL/Genbank database are shown in the fifth column. Presence or absence of telomere repeats, and TGAGGG, TTGGGG, TCAGGG or other variant repeats is shown in columns six and seven. The final column indicates whether a clone has been previously isolated and from which ordered array library.

chromosome ends are represented at least once (Yuri Dubrova, personal communication). This calculation is based on the assumption that all chromosome ends amplify equally well in the telomere-anchored PCR reaction, which is not an accurate statement. TVR telomere mapping at Xp/Yp (Baird *et al.*, 1995) and 12q (Baird *et al.*, 1999) showed that telomeres on homologous chromosomes can be allelic. Therefore the number of chromosome ends in the human genome is 92. For this calculation, an average library size of 752 clones is used, with 40 % clones positive for telomere repeats, giving a total 300 clones containing chromosome ends (excluding the possible presence of interstitial telomere-like repeat arrays). The probability that any one chromosome end is excluded is $(91/92)^{300}$. The probability that any one chromosome end is present is $[1 - (91/92)^{300}]$. It is rationalised that the probability that all 92 chromosome ends are found in 300 telomere positive clones is $[1-(91/92)^{300}]^{92} = 0.03$ (3 %). This is an upper estimate and will vary depending on the percentage of telomere positive clones in a library. This figure can be used to estimate the number of clones that would need to be picked in order to give 100 % probability that a library from one individual contains at least one copy of every chromosome end. Assuming a telomere-positive clone frequency of 40 %, an estimated 250,000 clones would have to be picked, which is beyond the practical limit of this strategy. However, these are theoretical values. It is perhaps more realistic to expect that most ordered array libraries contain a high proportion of chromosome ends, rather than all possible chromosome ends. The probability that this is true for most libraries is much higher. The greater the number of chromosome ends present the more likely it is that the breakpoint is included. Therefore the main aim of this work was to attempt to estimate the proportion of chromosome ends that were routinely efficiently amplified during telomere-anchored PCR and determine whether this could be improved within the technical limits of the strategy.

Size fractionation step of *Mbo*I-digested genomic DNA gives a 50-100 fold enrichment for long *Mbo*I fragments including telomeres. Cloning of PCR products amplified from these fragments can be very efficient, resulting in over 50 % telomere positive clones in some libraries (NT library 61 % and CEPH library 64 %, shown in table 4.1). However, this efficiency is often not reproducible. Low numbers of telomere positive clones lowers the probability of cloning and isolating a single chromosome end (such as a breakpoint). Comparison of the CEPH data with that obtained from the ordered array libraries described in chapter 3 is shown in table 4.2. The number of telomere positive clones is highly variable, ranging from 12 % to 64 %. Five of eight libraries contain less than 40 % telomere positive clones.

LIBRARY	TOTAL SIZE	TEL+		TSK37			TSK8			E-F			TSK46			CB0001-24c			CB0001-28		
		no.	%	no.	%	%+	no.	%	%+	no.	%	%+	no.	%	%+	no.	%	%+	no.	%	%+
FB336R	376	128	34	28	8	22	5	1.3	4		nd		6	2	5		nd			nd	
NT	376	230	61	37	10	16	5	1	2	31	8	14	7	2	3	8	2	4		nd	
FB241	376	74	20	5	1	7	1	0.3	1	15	4	20	3	1	4		nd			nd	
CB0001	329	140	43	19	6	14	1	0.3	0.7	31	9	22	4	1	3	3	1	2		nd	
CB	376	147	39	28	8	19	3	1	2	16	4	11	1	0.3	0.7	2	1	1.5	52	14	35
CB00054	752	210	28	24	3	11	1	0.1	0.5	18	2	9	0	0	0	2	0.3	1	80	11	38
ROBA	658	77	12	2	0.3	1	3	0.5	4	21	3	27	10	2	8	2	0.3	3	26	4	34
CEPH	1504	969	64	180	13	19	18	1	2	130	9	13	51	3	5	32	2	3	482	32	50

Table 4.2 The distribution of chromosome ends and subterminal repeat families in ordered array libraries

The library name is shown in the first column, indicating the patient DNA from which each library was generated. The second column indicates the total number of clones in each library. Column three shows the number and percentage of clones positive for telomere repeat arrays. Subterminal probes used for screening are indicated along the top row. For each subterminal probe, the number of positive clones are shown as follows: 1) the number of positives; 2) the number of positives as a percentage of the total number of clones in the library (%); and 3) the number of positives as a percentage of telomere positive clones (%+). nd indicates that screening with a particular probe was not carried out. Percentages are shown to the nearest whole number where appropriate. Data from libraries FB336R, NT and FB241 was supplied by N. J. Royle.

The distribution of some chromosome ends between libraries was compared, and indicated that telomere-anchored PCR consistently and efficiently amplified from a number of chromosome ends in independent reactions (shown in table 4.2). The Xp/Yp pseudoautosomal telomere-junction (Tsk8) was detected at a copy number of more than two in all libraries except FB241, CB0001 and CB00054. This chromosome end is therefore represented more than the expected number of times in most libraries. However, it cannot be assumed that this level of representation extends to all chromosome ends from each individual. The high copy number of the Xp/Yp telomere in most libraries is probably due to efficient amplification of this telomere-junction in the PCR step. The CEPH library would theoretically be expected to contain at least 30 copies of the Xp/Yp telomere, as it contains pooled DNA from 15 individuals. Only 18 copies were identified, although this was expected due to the difficulty in picking enough clones to ensure that all possible chromosome ends from a pool 15 individuals were represented, as previously discussed.

The 7q and 12q telomeres show a high level (97.9 %) of sequence similarity (Coleman *et al.*, 1999; Royle *et al.*, 1992) and the 7q and 12q telomere-junctions are detected by the E-F probe. Homologous telomere-junction clones Tsk48 and Tsk7 (Royle *et al.*, 1992) represent the 7q and 12q telomere-junctions respectively. Within the 2 kb of subterminal sequence adjacent to the 12q telomere, twenty-one polymorphic positions were found to be in complete linkage disequilibrium, creating three diverged haplotypes, 12qA, 12qA1 and 12qB, that account for 80 % of 12q telomeres in the Caucasian population (Baird *et al.*, 1999). The telomere-junction Tsk48 is homologous to the 12qA and 12qA1 haplotypes. The alignment of the telomere-junctions at 7q, 12qA, 12qA1 and 12qB is shown in figure 4.1. Given that a number of different structures exist at the 12q telomere, some individuals may have no copies, one copy, or two copies of the 12q telomere-junction detected by the E-F probe. It is possible that similar variation exists at the 7q telomere junction, and if so, the number of chromosome ends detected by the E-F probe could vary from 0-4 copies in different individuals. Two clones in the CB library positive for the E-F probe were sequenced and exhibited 15 % sequence divergence, suggesting that they had arisen from different chromosome ends. Clone CB-18g showed 95 % sequence identity to the 7q (Tsk48) telomere-junction, and clone CB-22g showed 100 % sequence identity to the 12qB allele. Figure 4.1 shows the alignment of the telomere-adjacent sequence from clone CB-18f with 7q and clone CB-22g with the three common 12q alleles (including Tsk7; accession no. X60162). This chromosome end was represented at high copy number in all individuals studied, and it is likely that clones arose from different chromosome ends in most libraries. However, because of the potential variation in the number

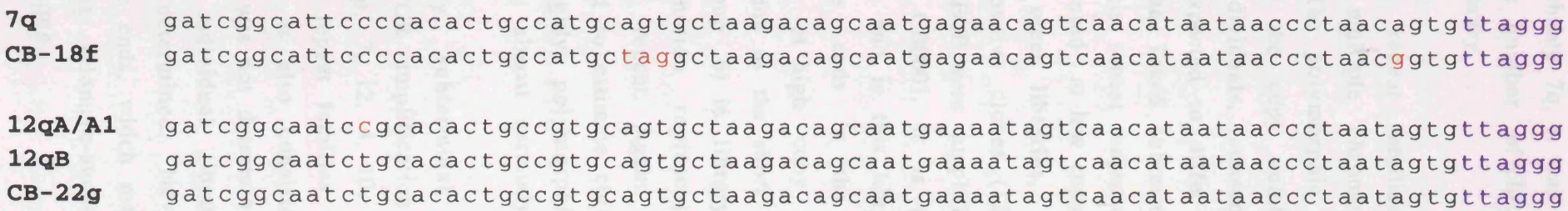


Figure 4.1 Alignment of telomere-adjacent sequence from clones CB-18f and CB-22g with the 7q and 12q telomere-junction sequences

Clones CB-18f and CB-22g were both positive for the E-F probe. These two clones show 15 % sequence divergence and were thought to represent isolates of different chromosome ends. The telomere-adjacent sequence from clone CB-18f shows 95 % sequence identity to the 7q telomere-junction clone (Tsk48). Variations from the 7q sequence are shown in red. The telomere-adjacent sequence from clone CB-22g shows 100 % sequence identity to the 12q haplotype B telomere. The Tsk7 telomere-junction clone is homologous to the 12qA and 12qA1 telomeres. The start of telomere repeats in each telomere-junction is shown in blue.

of copies of the telomere-junction detected by the E-F probe, due to the variation at 12q and probably 7q, it was not possible to use the number of E-F positive clones to estimate the number of chromosome ends represented by this telomere-junction in each library.

Subterminal repeat families such as Tsk46, Tsk37, PGSEGB and CB0001-24c are present at multiple chromosome ends, and are represented consistently in all libraries. The polymorphic distribution of many of these subterminal repeats means that the copy number of these repeats adjacent to a telomere varies between individuals, consequently the copy number of these telomere-junction clones is expected to differ between libraries. Library CB00054 did not contain any copies of the Tsk46 subterminal repeat, suggesting that this individual has no copies of this repeat adjacent to telomeres. The Tsk37 subterminal repeat family was represented at low copy number in the ROBA library (2 positives; 0.3 %) in contrast to other libraries, probably a direct consequence of the low number of telomere positive clones (12 %) in this library. Copies of the subterminal repeat family PGSEGB were amplified from at least ten different chromosome ends in individual CB0001. It is possible that there is full representation of this chromosome end in this library, as the PGSEGB repeat amplified from at least four chromosome ends in the monochromosome hybrid panel. This repeat was represented at high copy number in all libraries characterised, and amplifies efficiently due to the short stretch of telomere-adjacent sequence. Similarly, four clones (group 6) in library CB0001 containing the PGB4G7 subterminal repeat telomere-junction represented at least three different alleles or isolates of this subterminal repeat. Again, the copy numbers of these subterminal repeats could not be used to estimate chromosome end distribution in ordered array libraries due to their highly polymorphic nature, as their copy number and chromosomal location are almost certainly dependent on the genotype of each individual.

In summary, subterminal probe hybridisation data indicates that telomere-anchored PCR amplification consistently and efficiently occurs from chromosome ends Xp/Yp, 7, 12, 4, 10, 11, 17, 18, 20, 17. Amplification from polymorphic subterminal repeat families on a number of different chromosomes, such as Tsk37 and Tsk46, is also consistent, although their chromosomal location in each individual was not determined. Novel telomere-junction clones were also isolated from most individual libraries, although their chromosomal location could not always be determined. Nevertheless, these clones represent amplification from chromosome ends, which may or may not have been previously isolated. This data suggests that telomere-anchored PCR amplification is optimal, within the limits of the technique.

Section 2. Long-PCR

4.4 Optimisation of long-PCR

Long-PCR amplification was used to increase the efficiency of amplifying long *Mbo*I-digestion products. These products potentially contain terminal deletion breakpoint fragments where the most distal *Mbo*I site at the chromosome terminus lies some distance from the telomere. Long-PCR conditions were optimised as described in chapter 2 (Materials & Methods – DNA amplification), and assayed using NT genomic DNA. NT has a 130 kb terminal deletion at 22q13.3, and a phenotype of mild mental retardation and delay in expressive speech. The NT breakpoint lies within minisatellite locus MS607 (Wong *et al.*, 1997). MS607 consists of two discrete repeat arrays, MS607A and MS607B (Armour and Jeffreys, 1991), estimated to be no more than 1 kb apart. The breakpoint is located 892 bp distal to MS607A (Wong *et al.*, 1997) and therefore lies at the 5′ end, or within minisatellite MS607B. The paternal MS607B allele is missing on the truncated chromosome. The structure of MS607 and the location of the breakpoint is shown in figure 4.2. MS607A has limited variability (around 50 % heterozygosity), and alleles range from 2.3 kb to approximately 4 kb. The most proximal *Mbo*I site lies about 277 bp from the start of MS607A, and the NT *Mbo*I breakpoint fragment, including the MS607A allele, is 3.5 kb in length (N. J. Royle, personal communication).

Long-PCR amplification of the MS607A locus with flanking primers (MS607A and MS607B) in NT and unrelated individuals is shown in figure 4.3. Cycling conditions were 95°C for 1 min, 67°C for 1 min and 70°C or 10 min for 32 cycles, followed by a 'chase' cycle of 67°C for 1 min and 70°C for 10 min to minimise amplification of products shorter than the target sequence (Armour and Jeffreys, 1991). Southern hybridisation with probe MS607A detected a common product of approximately 3.5 kb in patient NT, and unrelated individuals P50, P53, P55 and DW. Individual P55 also showed amplification of a second product of approximately 3.8 kb, and individual P50 a second product of 3.7 kb, representing allele heterozygosity in these individuals.

The MS607A locus lies immediately adjacent to an array of telomere repeats at the NT breakpoint. Amplification of this locus immediately adjacent to an array of telomere repeats in NT was detected by PCR amplification with primer MS607A and telomere primer TelC, followed by Southern hybridisation with the MS607 probe. Long-PCR conditions resulted in significant enrichment for the 3.5 kb breakpoint fragment in NT. Amplification under normal PCR conditions was detected at only very low levels after long autoradiograph exposure times (5-7 days), whereas

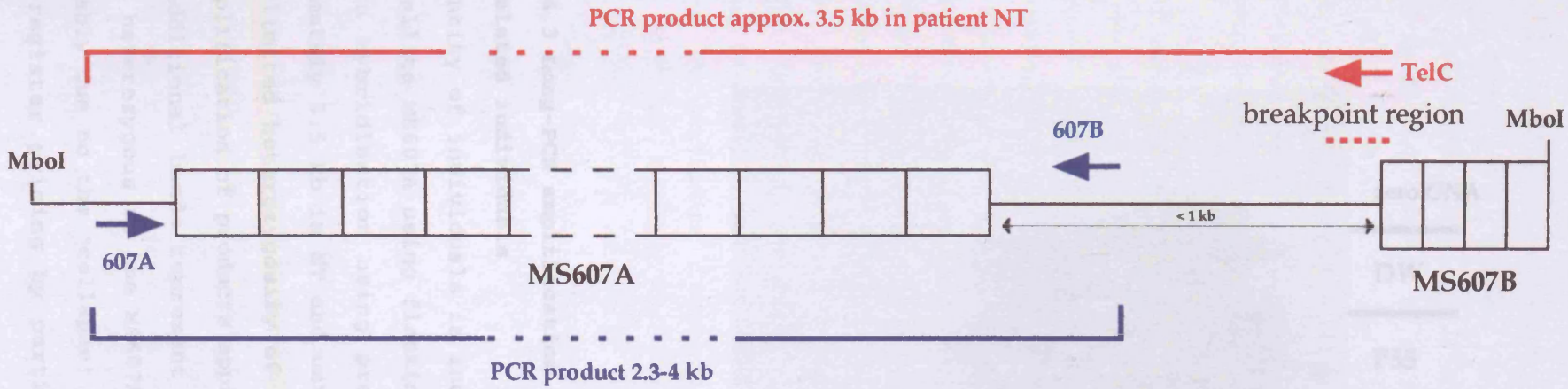


Figure 4.2 Structure of minisatellite locus MS607 (D22S163), and location of the NT breakpoint

Minisatellite locus MS607 (D22S163) consists of two discrete minisatellite repeats arrays, MS607A and MS607B (Armour and Jeffreys, 1991). MS607A has limited variability, with allele sizes ranging from 2.3 kb to approximately 4 kb. MS607B accounts for most of the variability at MS607. MS607A and MS607B are estimated to be no more than 1 kb apart. The most proximal MboI site in this region lies 227 bp from the start of the MS607A repeat array. The NT breakpoint lies 892 bp distal to MS607A (Wong *et al.*, 1997), and is therefore located at the 5' end, or at the start, of the MS607B repeat array. The breakpoint region is indicated by the red dashed line. MS607A flanking primers (MS607A and MS607B) are indicated by blue arrows, and their PCR amplification product indicated in blue. Telomere primer TelC is indicated by the red arrow, and the PCR product generated by amplification of the NT breakpoint fragment with primers MS607A and TelC is indicated in red.

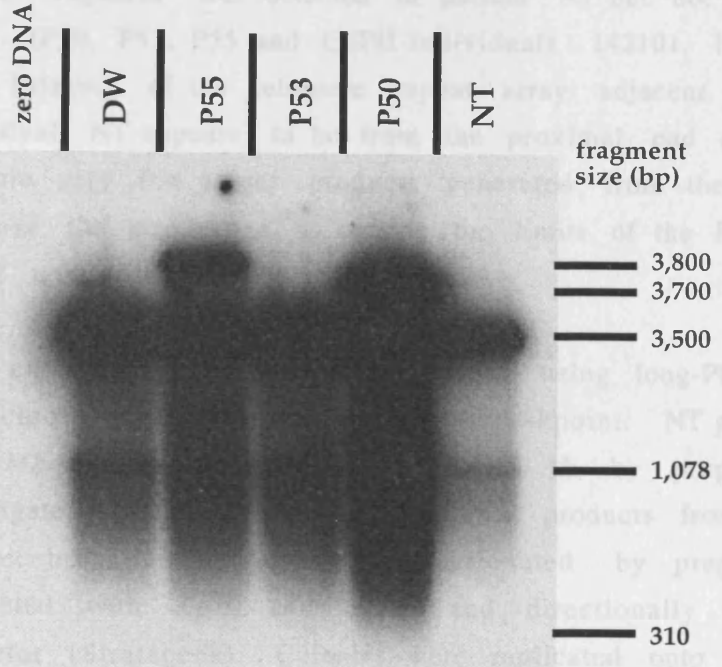


Figure 4.3 Long-PCR amplification of the MS607 minisatellite locus in NT and unrelated individuals

The identity of individuals is indicated above each lane. Amplification of minisatellite MS607A using flanking primers under long-PCR conditions, and Southern hybridisation using probe MS607 detected a common product of approximately 3.5 kb in NT and unrelated individuals P50, P53, P55 and DW, due to limited heterozygosity of this locus. Individuals P55 and P50 also show amplification of products approximately 3.8 kb and 3.7 kb respectively. These additional bands represent other alleles in these two individuals, who are heterozygous at the MS607A minisatellite locus. The extended smear is probably due to the 'collapse' of products from the minisatellite due to out-of-register priming by partially extended products (Armour, 1990).

long-PCR conditions enabled detection of the breakpoint fragment after normal exposure times (18-24 hours), shown in figure 4.4, indicating a significant increase in amplification efficiency under long-PCR conditions. Amplification of the 3.5 kb breakpoint fragment was detected in patient NT but not in six other unrelated individuals (P50, P53, P55 and CEPH individuals 142101, 133401, 0201), shown in figure 4.5. Priming of the telomere repeat array adjacent to MS607 by primer TelC in individual NT appears to be from the proximal end of the repeat array only, as there are very few larger products generated from the breakpoint. This is perhaps because the size range is outside the limits of the PCR reaction, even under long-PCR conditions.

Due to the successful enrichment for the NT breakpoint using long-PCR, a second attempt was made to clone the NT terminal deletion breakpoint. NT genomic DNA was digested with *MboI*, size fractionated above 5 kb by preparative gel electrophoresis, and ligated to Sau3AI-linkers. Long-PCR products from 3 kb to 4.3 kb encompassing the breakpoint, were size fractionated by preparative gel electrophoresis, digested with *Kpn*I and *Eco*RI and directionally cloned into pBluescriptII SK+ vector (Stratagene). Colonies were replicated onto nylon filters and screened with the MS607 probe to detect recombinant clones containing the breakpoint (data not shown). No positive colonies were detected, although there was a high density of recombinant clones on agar plates. It is likely that the breakpoint fragment was not cloned due to its size (3.5 kb), or that clones containing the breakpoint fragment were highly unstable due to their internal structure. GC-rich minisatellites have been shown to display poor stability in plasmid vectors. This is consistent with previous analysis of a NT ordered array library, where breakpoint fragments were detected in telomere-anchored PCR products at low copy number but were not cloned (N. J. Royle, personal communication).

4.5 Constructing libraries using long-PCR

Long-PCR resulted in significant enrichment for the NT breakpoint fragment in comparison to normal PCR conditions. Consequently, long-PCR was used to generate ordered array libraries from patients CB and CB00054 (described in chapter 3), AJ, NS and BQ, in order to increase the probability of amplifying breakpoints that may be located in a long *Mbo*I fragment. AJ and NS have deletions at 22q13.3, and a phenotype including developmental delay and mild facial dysmorphia (Nesslinger *et al.*, 1994). BQ has a suspected terminal deletion at 4p (gift from Dr. J. Flint). However, screening libraries generated from AJ, BQ and NS with the Tel+ telomere repeat probe detected only one positive clone in AJ, and no

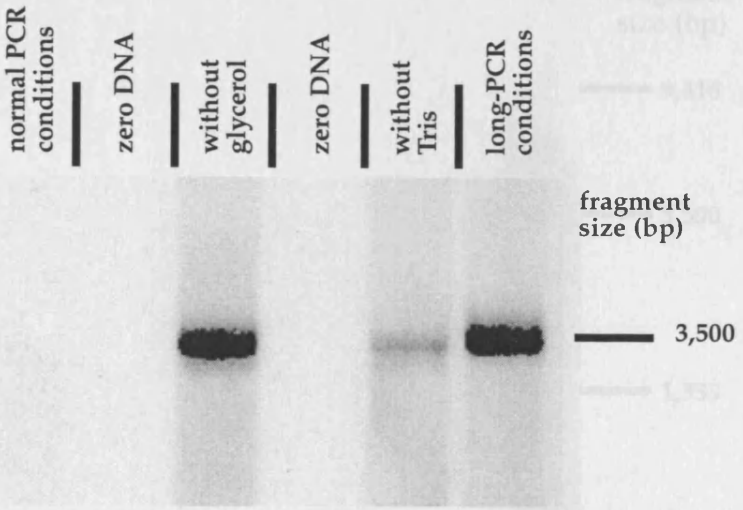


Figure 4.4 Increasing the amplification efficiency of the NT breakpoint using long-PCR conditions

Amplification of the NT breakpoint fragment was carried out using primer MS607A and telomere primer TelC, and detected by Southern hybridisation using probe MS607A. Comparison of products amplified under normal and long-PCR conditions indicated that long-PCR resulted in a significant improvement in amplification efficiency of the breakpoint. Glycerol can be added to enable DNA strands to separate at lower temperatures, however, it did not influence the amplification of the NT breakpoint and was not used in subsequent reactions. Amplification of the NT breakpoint fragment under normal conditions was detected at only very low levels even after long autoradiograph exposure times (1-2 weeks). Detection of the breakpoint product under long-PCR conditions was possible after normal exposure times (18-24 hours).

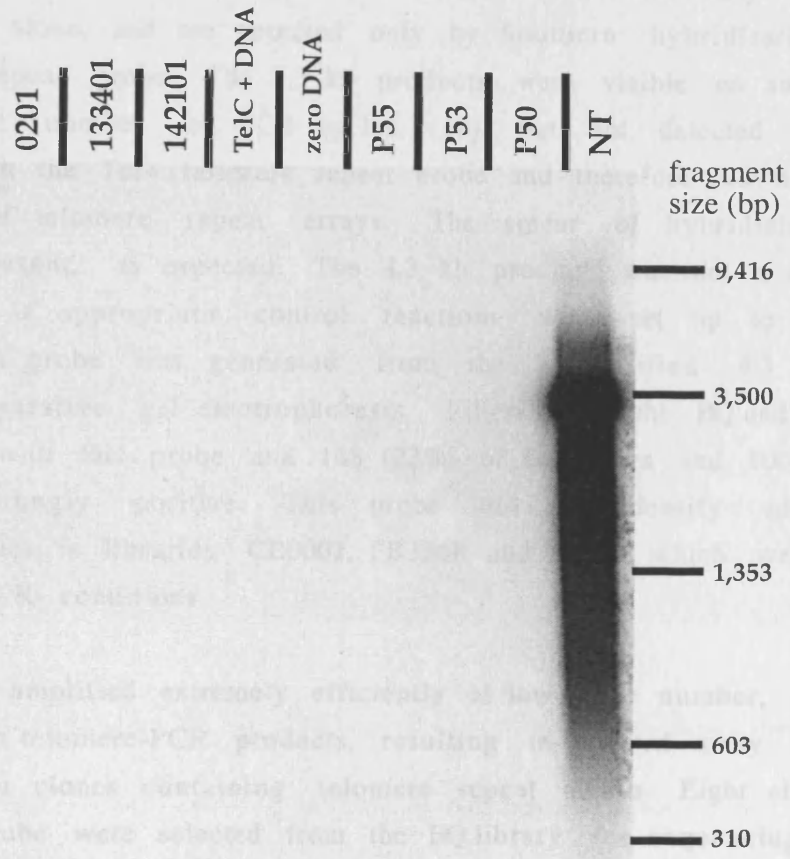


Figure 4.5 Long-PCR amplification of the NT breakpoint

The identity of DNAs used are indicated above each lane. Long-PCR amplification using primer MS607A and telomere primer TelC, and Southern hybridisation using probe MS607A detected amplification of the expected 3.5 kb breakpoint product in individual NT only. Other unrelated individuals did not show amplification of this product due to the absence of telomere repeats adjacent to the MS607A locus in these individuals. The extended smear seen in NT is most likely due to the collapse of products from the minisatellite repeat array due to out-of register priming by partially extended products (Armour, 1990).

positive clones in BQ and NS. It was established that long-PCR conditions had resulted in the novel generation of a population of fragments, approximately 4.3 kb in size. Cycle numbers used for telomere-anchored PCR are normally kept to a minimum (generally less than 20 cycles). Products are not normally visible by gel electrophoresis alone, and are detected only by Southern hybridisation with the Tel+ telomere repeat probe. The 4.3 kb products were visible on an agarose gel after a minimal number of PCR cycles (18), but not detected by Southern hybridisation with the Tel+ telomere repeat probe and therefore did not arise from amplification of telomere repeat arrays. The smear of hybridising telomere products was present, as expected. The 4.3 kb product was not a result of PCR contamination, as appropriate control reactions were set up to monitor and prevent this. A probe was generated from the unidentified 4.3 kb long-PCR products by preparative gel electrophoresis. Filters from the BQ and NS libraries were screened with this probe and 168 (22%) of BQ clones and 100 (13%) of NS clones were strongly positive. This probe did not identify any positively hybridising clones in libraries CB0001, FB336R and CEPH, which were generated under normal PCR conditions.

As this product amplified extremely efficiently at low cycle number, it was cloned in preference to telomere-PCR products, resulting in ordered array libraries that were negative for clones containing telomere repeat arrays. Eight clones positive for the 4.3 kb probe were selected from the BQ library for sequencing to identify the insert. However, the clones did not show homology to each other and had variable insert sizes of less than 1.3 kb, indicating that this product arises from a population of molecules rather than a single molecule. Clone 1 showed 88 % identity to an *Alu* element from the Jo-subfamily, and clone 3 showed 94 % identity to an *Alu* element from the Sc-subfamily. Clone 4 showed 97 % identity to an L1 repeat. Clone 7 showed 94 % identity with a region of chromosome 1q24-25, and clone 2 showed 91 % identity with a region of 19q13.2, but no further information was available on these sequences. Clones 5, 6 and 8 did not show homology with any sequences available in the EMBL/Genbank databases. These clones do not contain common sequences, although three contain different interspersed repetitive elements. *Alu* repeats are approximately 330 bp in length (Batzer *et al.*, 1993), and are therefore unlikely to directly give rise to amplification of this 4.3 kb product. It was first thought that this population of PCR products represented an abundant repeated region resistant to *Mbo*I digestion, which was included in the size fractionation step. However, the 4.3 kb PCR products were completely digested with *Mbo*I. Further investigation indicated that amplification of these 4.3 kb products was dependent on the presence of both the Sau3AI-linker and the SauL-A linker primer. They did not amplify from undigested genomic DNA, even

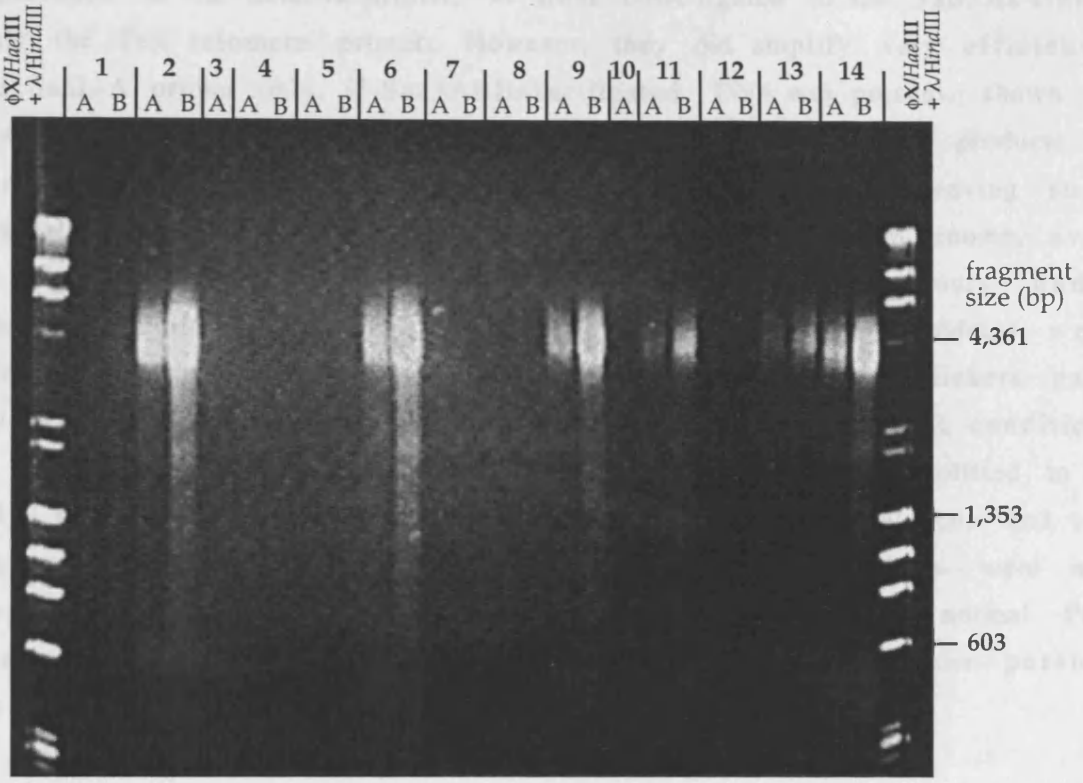


Figure 4.6 Amplification of an unidentified 4.3 kb fragment under long-PCR conditions

Long-PCR conditions were used to generate telomere-anchored PCR products for cloning from patients CB and NS. This resulted in the efficient amplification of a heterogenous population of products with an average length of 4.3 kb, visible on an agarose gel with ethidium bromide staining. These products did not hybridise to the telomere repeat probe. Amplification of the 4.3 kb product is dependent on the presence of the Sau3AI-linker and the linker primer SauL-A (SLA). It is likely to be a result of only partial MboI digestion of high molecular weight genomic DNA, resulting in the presence of large non-telomeric fragments which are included in the size fractionation and ligated to the Sau3AI-linker. Amplification of these products is efficient due to changes in the reaction conditions used for long-PCR amplification.

Key

A - 18 cycles of PCR amplification
B - 20 cycles of PCR amplification

1 - CB genomic DNA + SLA + TelC
2 - CB with linker + SLA
3 - CB with linker + TelC
4 - SLA only
5 - SLA only
6 - CB with linker + SLA + TelC
7 - No DNA + SLA + TelC
8 - NS genomic DNA + SLA + TelC
9 - NS with linker + SLA
10 - NS with linker + TelC
11 - NS with linker + TelC
12 - No DNA + SLA + TelC
13 - NS with linker + SLA + TelC
14 - CB with linker + SLA + TelC

in the presence of the SauL-A primer, or from DNA ligated to the Sau3AI-linker with only the TelC telomere primer. However, they did amplify very efficiently with the SauL-A primer only, if Sau3AI-linker-ligated DNA was present, shown in figure 4.6. The most likely explanation for presence of these 4.3 kb products is incomplete *Mbo*I digestion of high molecular weight genomic DNA, leaving some long, partially digested products from non-telomeric regions of the genome, even though digestion reactions were incubated for at least 2 hours under recommended reaction conditions. These long, partially digested products were size selected with telomere fragments. Subsequent ligation of Sau3AI-linkers gave rise to long PCR fragments with linkers ligated at each end. Long-PCR conditions provide optimum amplification for products of this size, which are amplified to an unusually high copy number and then cloned due to their prevalence, and the presence of internal *Eco*RI and *Kpn*I sites. These 4.3 kb products were not amplified during generation of other ordered array libraries as normal PCR conditions were used. Consequently, long-PCR amplification was not pursued further.

Section 3. A pre-cloning size fractionation step

4.6 Size fractionation of PCR products prior to cloning

Long-PCR amplification was investigated in order to increase the efficiency of amplifying long terminal deletion breakpoint fragments. However, cloning efficiency of long PCR products is inefficient due to preferential cloning of smaller, high copy number products. Size fractionation of PCR products prior to cloning, to isolate a population of longer PCR products, would reduce competition and lead to an increase in the efficiency of cloning these products. To assess the value of a pre-cloning size fractionation step, two ordered array libraries were generated with genomic DNA from patient ROBA with a deletion at 1p36.3 (Giraudeau *et al.*, 1997), previously described in Chapter 3. Before cloning, telomere-anchored PCR products were separated into two size fractions which were cloned separately, generating two libraries. Clone structure and sequence data from these libraries was then compared to determine whether size fractionation of telomere-anchored PCR products resulted in an increased success rate in the cloning of longer PCR products, and consequently a larger subset of chromosome ends.

ROBA genomic DNA was prepared for telomere-anchored PCR as previously described. After PCR amplification with the SauL-A linker primer and telomere primer TelC under normal PCR conditions, PCR products were size fractionated by

preparative agarose gel electrophoresis. Two fractions were isolated; fraction one comprised PCR fragments from 310 bp to 1.5 kb, and fraction two comprised fragments from 1.5 kb to approximately 10 kb. The fractions were cloned separately to generate two libraries - ROBAIIA (310 bp - 1.5 kb) and ROBAIIB (1.5 kb - 10 kb).

4.7 Characterisation of ROBAIIA clones

The library contained a total number of 752 clones. 34 % (258) were positive for the Tel+ telomere repeat probe. After screening with subterminal probes (data shown in Appendix I), 61 (24 %) telomere positive clones remained to be sequenced. All clones contained the predicted sequence from the *Kpn*I to *Sau*3A sites of the Sau3AI-linker unless otherwise stated. ROBAIIA clones sequenced are summarised in table 4.3. Clones **RIIA-18c** and **RIIA-42e** did not show significant homology to sequence in the EMBL/Genbank databases. However, both clones contained variant repeats in the telomere repeat array and were therefore not investigated further. These clones probably represent novel telomeres from normal chromosome ends. Clone **RIIA-51d** showed sequence identity to an *Alu* element of the Sx-subfamily. However, telomere repeats were not detected on sequencing. This clone was weakly positive with the Tel+ telomere repeat probe and this may be due to the presence of a second 'unstable' colony in the well containing telomere repeats.

4.8 Characterisation of ROBAIIB clones

Library ROBAIIB was generated by cloning telomere-anchored PCR products larger than 1.5 kb. It was expected that the cloning efficiency would be lower due to the size of the PCR products. The library contained a total number of 752 clones, but only 3 % (20) were positive with the Tel+ telomere repeat probe. It was not clear why the number of clones containing telomere repeats was so low when only recombinant colonies were picked. The presence of telomere positive clones in the ROBAIIA indicated that PCR amplification of digestion products containing telomere repeat arrays was efficient. It is possible that clones containing very long arrays of telomere repeats are unstable, undergoing rearrangement and loss of insert at high frequency. Similar instability has been observed in GC-rich minisatellites when cloned into plasmid vectors (N. J. Royle, personal communication). The ROBAIIB library was screened with the Tsk8 (Xp/Yp pseudoautosomal telomere-junction) probe only, for which two clones were positive. All other telomere positive clones were sequenced and contained the predicted sequence from the *Kpn*I to *Sau*3A sites of the Sau3AI-linker, unless

Clone name	No. of clones in group	Start site	Telomere-adjacent sequence	Sequence indentity FASTA/BLAST database search	Telomere repeats present	Variant repeats present	No. of repeat interspersion patterns in group	Previously identified?
GROUP 1	5	*Kpn* I	no	_	yes	no		_
GROUP 2	32	*Sau* 3A	no	_	yes	TGAGGG, TCAGGG, TTGGGG other variants		_
GROUP 3	2	*Sau* 3A	16 bp	_	yes	TCAGGG other variants	1	_
RIIA-19b		*Sau* 3A	46 bp	96 % subterminal repeat (U53226)	yes	TGAGGG, TCAGGG other variants		CB
RIIA-29h		*Sau* 3A	75 bp	88 % subterminal-like repeat (X64633)	yes	other variant repeats		no
RIIA-18c		*Sau* 3A	102 bp	no	yes (7)	2 variant repeats		_
RIIA-42e		*Sau* 3A	61 bp	no	yes	variant repeats		_
RIIA-51d		*Sau* 3A	> 200	93 % Alu-Sx	no			_

Table 4.3 Summary of clones sequenced from the ROBAIIA library

Clone names are shown in the first column. Clones are grouped according to similarity either in the telomere-adjacent DNA, or telomere repeat array if adjacent DNA is absent. The number of clones in each group, and the start site, Sau3A (GATC) or KpnI (GGTACC), are shown in columns two and three. The length of telomere-adjacent sequence in each group of clones is shown in basepairs. Results of sequence homology searches using BLAST and FASTA searches at the EMBL/Genbank database are shown in the fifth column. Presence or absence of telomere repeats, TGAGGG, TTGGGG, TCAGGG or other variant repeats is shown in columns six and seven. The number of interspersion patterns within the telomere repeat arrays of each group of clones is indicated in column eight. The final column indicates whether a clone has been previously isolated and from which ordered array library.

otherwise stated. A summary of clones sequenced from the ROBAIIB library is shown in table 4.4.

Clone RIIB-18d had 210 bp of telomere-adjacent sequence with 99 % homology to the 7q telomere (AF027390) including the telomere junction (Tsk48), and would have been detected by probe E-F. This clone clearly contained the expected *Mbo*I site at which the 7q telomere-junction is normally cleaved, generating a telomere with 74 bp of adjacent sequence. However, this amplicon was not cut at this site but at a more internal *Mbo*I site, generating 210bp of telomere-adjacent sequence, and may result from partial digestion of genomic DNA. This clone also had an array of telomere repeats including TCAGGG, TTGGGG and other variant repeats.

Clone RIIB-66c had 450 bp of telomere-adjacent sequence without significant homology to sequences in the EMBL/Genbank databases and an array of telomere repeats with variants. Therefore this clone was not investigated as a candidate breakpoint and probably represents a novel chromosome end or a polymorphic telomere.

Group 5. Two identical clones had 95 bp of telomere-adjacent sequence with 89 % homology to a telomere-associated repeat found at 2p11 (AC002038). This clone had an array of telomere repeats including variants (TCAGGG) and many other degenerate repeats, a feature of interstitial telomere-like repeat arrays (Wells *et al.*, 1990).

Group 6 clones did not show any identity to sequences in the EMBL/Genbank database, and telomere repeats were not evident on the sequencing gel. These clones showed only very weak hybridisation signals with the telomere repeat probe and probably represent false positives.

4.9 Distribution of chromosome ends in ROBAIIA and ROBAIIB

The insert sizes of all ROBAIIB sequenced clones were determined by *Eco*RI and *Kpn*I digestion, and ranged from approximately 1.5 kb to 2.3 kb. This size range was clearly larger than normally observed in telomere-anchored PCR libraries, which ranges from 600 bp to 1.5 kb, but is generally less than 1 kb. This demonstrated than the pre-cloning size fractionation step resulted in the cloning of longer PCR products. Analysis of ROBAIIB clones showed that most contained PCR products with longer telomere repeat arrays but the telomere-adjacent DNA was less than 450 bp in all clones, which is regularly achieved using the cloning step without size fractionation. In addition, the efficiency of cloning telomere-repeat arrays was very low (3 %), ruling out size fractionation as a useful additional step.

Clone name	No. of clones in group	Start site	Telomere-adjacent sequence	Sequence indentity FASTA/BLAST database search	Telomere repeats present	Variant repeats present	No. of repeat interspersion patterns in group	Previously identified?
GROUP 1	1	*KpnI*	no	–	yes	no	–	–
GROUP 2	3	*Sau3A*	no	–	yes	TCAGGG other variants	–	–
GROUP 3	3	*Sau3A*	28 bp	PGSEGB	yes	variant repeats	2	CB0001
GROUP 4	3	*Sau3A*	300 bp	98 % subterminal repeat (X58156)	yes	variant repeats	1	
GROUP 5	2	*Sau3A*	95 bp	89 % telomere-associated repeat chromosome 2	yes	TCAGGG degenerate repeats	1	
GROUP 6	3	*Sau3A*	> 100 bp	no	no	–	–	
RIIB-19f		*Sau3A*	73 bp	E-F	yes	TTGGGG other variants	–	
RIIB-18d		*Sau3A*	210 bp	E-F	yes	TCAGGG, TTGGGG other variants	–	
RIIB-66c		*Sau3A*	450 bp	no	yes	variant repeats	–	

Table 4.4 Summary of clones sequenced from the ROBA-IIB library

Clone names are shown in the far left-hand column. Clones are grouped according to similarity either in the telomere-adjacent DNA, or telomere repeat array if adjacent DNA is absent. The number of clones in each group, and the start site, *Sau* 3A (GATC) or *Kpn* I (GGTACC), are shown in columns two and three. The length of telomere-adjacent sequence in each group of clones is shown in basepairs. Results of sequence homology searches using BLAST and FASTA search tools at the EMBL/Genbank database are shown in the fifth column. Presence or absence of telomere repeats, and TGAGGG, TTGGGG, TCAGGG or other variant repeats is shown in columns six and seven. The number of interspersion patterns within the telomere repeat arrays of each group of clones is indicated in column eight. The far right-hand column indicates whether a clone has been previously isolated and from which ordered array library.

Due to the low number of telomere positive clones in the ROBAIIB library, it was not possible to draw any firm conclusion as to whether size fractionation of PCR products prior to cloning influences the distribution of chromosome ends present in a library. Both ROBAIIA and ROBAIIB contained a high proportion of telomere positive clones with telomere repeat arrays only, and clones containing the Xp/Yp pseudoautosomal telomere junction (Tsk8) and the 7q/12q telomere-junction (E-F). However, there were some differences. The ROBAIIB library did not contain copies of the Tsk37, Tsk46 or CB0001-24c subterminal repeat families normally found on a number of chromosome ends. These repeats were present in the ROBAIIA library and the previously described ROBA library, and were therefore amplified efficiently in the PCR step but excluded during size fractionation or cloning of ROBAIIB. Although the exclusion of these ends is not important per se, as they are normally excluded by screening with probes prior to sequencing, it strongly suggests that other ends may also have been excluded, including the breakpoint. It is likely that these repeats were not found in the ROBAIIB library as a direct consequence of the low numbers of telomere positive clones. Subterminal repeats U53226 and X64633 were isolated from the ROBAIIA library only. These subterminal repeats were previously isolated from other patient libraries but not the primary ROBA library. The subterminal repeat X58156 (previously isolated from the CB library), and telomere-associated repeat AC002038, were isolated from the ROBAII-B library only, shown in table 4.5.

The ROBAIIA library was later screened with variant repeat probes TelG (TGAGGG), TelJ (TTGGGG) and TelK (TCAGGG) to investigate the distribution of these variant repeat types. The number of telomere clones positive for variants repeats TGAGGG, TTGGGG and TCAGGG were 160 (62%), 139 (54%) and 24 (9 %) respectively. The ROBAIIB library was not screened with variant repeat probes due to the low number of telomere positive clones. However, sequence analysis revealed that 12 (63 %) clones contained TCAGGG repeats and 2 (11 %) clones contained TTGGGG repeats. No clones containing TGAGGG repeat were sequenced. The distribution of repeat types was very different between libraries. TCAGGG variant repeats were common in ROBAIIB, but found in low numbers in ROBAIIA, and TGAGGG repeats were common in ROBAIIA but not present at all in ROBAIIB. It is not clear why these differences in repeat distribution exist, although the low number of telomere positive clones in ROBAIIB may have introduced a bias in the population of telomeres present. Alternatively, PCR amplification from the proximal ends of much longer template arrays may lead to misrepresentation of the distribution of variant repeat types within a particular telomere repeat array in ROBAIIA. Telomere mapping at the Xp/Yp pseudoautosomal telomere (Baird *et al.*, 1995), the 7q, 12q and 16p telomeres (Baird *et al.*, 1999; Coleman, 1998, Coleman *et al.*, 1999)

Table 4.5 Comparison of clones in size fractionated libraries ROBAIIA and ROBAIIB

Library ROBAIIA was generated using telomere-anchored PCR products from 300 bp-1.5 kb, and ROBAIIB generated using PCR products from 1.5 kb to approximately 10 kb. Sequence identities derived from the EMBL/Genbank database using FASTA and BLAST searches are indicated in the left-hand column. The number (and percentage) of clones positive for each sequence in these two libraries are indicated. Some clones in the ROBAIIA library were positive for more than one subterminal repeat sequence.

	ROBAIIA	ROBAIIB
Telomere positive clones	258 (34 %)	20 (3 %)
Telomere repeats Only	39 (15 %)	4 (20 %)
Xp/Yp telomere junction (Tsk8)	6 (2 %)	2 (10 %)
PGSEGB subterminal Repeat (CB0001-28)	125 (48 %)	3 (15 %)
7q/12q telomere junction (E-F)	58 (23 %)	2 (10 %)
Tsk37 subterminal repeat family	72 (24 %)	-
Tsk46 subterminal repeat family	15 (6 %)	-
CB0001-24 subterminal repeat family	12 (4 %)	-
Subterminal repeat (mainly acrocentric chromosomes; U53226)	1 (0.5 %)	-
Subterminal-like repeat (X64633)	1 (0.5 %)	-
HST subterminal repeat (X58156) (Cheng *et al.*, 1991)	-	3 (15 %)
Telomere-associated repeat found on chromosome 2 (AC002038)	-	2 (10 %)
Clones with telomere repeats not identified by database searches	2 (1 %)	1 (5 %)
Clones without telomere repeats	1 (0.5 %)	3 (15 %)

indicated hypervariable repeat interspersion patterns. In general, TCAGGG repeats tend to be confined to the first 100 repeats in these telomeres, whereas complex interspersion patterns of TGAGGG repeats can extend at least 200 repeats into the telomere repeat array. In the Xp/Yp telomere most of the variation stops at around 120 repeat units into the array (Baird, 1996,), and no distal limit of variation has yet been identified at the 7q and 12q telomeres. It is possible that differences in repeat distribution patterns exist in other telomeres, and this may explain the varying distribution of variant repeats between the ROBAIIA and ROBAIIB libraries.

Discussion

4.10 Distribution of chromosome ends in ordered array libraries

Work described in this chapter was carried out in order to investigate and optimise the telomere-anchored PCR and cloning steps in this strategy. The main focus of this work was to determine the number of chromosome ends consistently amplified in the telomere-anchored PCR step, and whether this needed to be improved. The CEPH ordered array library was generated to examine the distribution of normal chromosome ends amplified from genomic DNA in the telomere-anchored PCR step, by maximising the number of target chromosome ends in the reaction. It was also expected the high proportion of target chromosome ends in the initial population of molecules would yield a number of novel chromosome ends or polymorphic telomeres, although no such ends were identified. Hybridisation and sequence data from the CEPH library was compared to data obtained from previously characterised patient libraries to look at the distribution of chromosome ends in each library, and to ascertain whether telomere-anchored PCR gave consistent amplification of a high proportion of chromosome ends. Due to the limitations of the strategy, it was not possible to pick enough clones containing telomere repeats to ensure that libraries have 100 % probability of containing at least one copy of all available chromosome ends (92 in total). However, telomere-anchored PCR was shown to consistently amplify from chromosomes Xp/Yp, 4, 7, 10, 11, 12, 17, 18, 20, and also high copy numbers of subterminal repeat families Tsk37, Tsk46, PGB4G7 and PGSEGB, found adjacent to telomeres on more than one chromosome end. Clones within a library containing identical copies of subterminal repeat sequence families often displayed different repeat interspersion patterns in the telomere repeat array, suggesting that these clones represent different alleles or copies from different chromosome ends. Additional chromosome ends are often isolated during library sequencing, although their specific chromosomal location cannot always be identified. It should be noted that screening patient libraries with variant repeat probes also led to the exclusion of a large number of normal chromosome ends prior to sequencing. In summary, comparison of hybridisation and sequence data from all libraries generated to date indicated that amplification of a representative number of chromosome ends consistently occurring in individual PCR telomere-anchored reactions.

4.11 Long-PCR amplification

Long-PCR resulted in significant enrichment for the 3.5 kb NT *Mbo*I breakpoint fragment, which is not efficiently amplified under normal PCR reaction conditions. However, attempts to clone the NT breakpoint fragment were unsuccessful, probably due to its size or unstable internal structure. Long-PCR was used to generate ordered array libraries from patients CB, CB00054, BQ, AJ and NS. However, long-PCR conditions introduced a novel problem - the efficient amplification of a heterogenous set of non-telomere products with a size of approximately 4.3 kb. These products were cloned preferentially to telomere-repeat containing fragments due to their high copy number, resulting in libraries lacking telomere positive clones. The mixed 4.3 kb product probably arose from partial *Mbo*I digestion of high molecular weight genomic DNA, allowing non-telomere sequences to get through the size fractionation step. It is likely that only a random, very small proportion of molecules fail to digest completely, but long-PCR conditions allow very efficient amplification of these non-telomeric fragments, resulting in high copy numbers from a small initial population. As a result, long-PCR could not be used as an effective modification to the general strategy.

4.12 Size fractionation of PCR products prior to cloning

Short PCR products are cloned much more efficiently than longer products. Cloning of short products may introduce bias into the distribution of chromosome ends in a library, by preferentially selecting telomere fragments with short regions of telomere-adjacent DNA. Size fractionation of telomere-anchored PCR products prior to cloning was investigated with a view to improve cloning efficiency of longer telomere-anchored PCR products by reducing competition from smaller PCR products. Comparison of telomere-junction clones from libraries ROBAIIA (size fraction 300 bp to 1.5 kb) and ROBAIIB (size fraction 1.5 kb to approximately 10 kb) indicated that size fractionation resulted in an increase in insert size of clones. However, this was a result of cloning fragments with longer telomere repeat arrays rather than telomere adjacent DNA. The efficiency of cloning telomere repeat arrays was still very low (3 %), consequently size fractionation and cloning of larger PCR products was not a useful modification to the general strategy. The low number of telomere positive clones in the ROBAIIB library meant that it was not possible to determine whether insert size influenced the distribution of chromosome ends in comparison to the ROBAIIA library.

Attempts to modify the PCR and cloning steps used to construct ordered array libraries were unsuccessful. It was thought that the most effective way to improve the efficiency of the strategy would be the introduction of an additional enrichment for telomere repeat arrays, to use in conjunction with *Mbo*I-digestion and size fractionation, and telomere-anchored PCR. The development of this additional step is described in chapter 5.

Chapter 5

A FILTER HYBRIDISATION SELECTION STEP TO ENRICH FOR TELOMERE REPEAT ARRAYS IN ORDERED ARRAY LIBRARIES

Summary

Cloning of telomere repeat arrays using *Mbo*I digestion, size fractionation and telomere-anchored PCR can be efficient, but often highly variable. An additional telomere enrichment step, based on a filter hybridisation method, was introduced, and shown to be very effective, particularly for telomere repeat arrays containing TTAGGG repeats only, including a terminal deletion breakpoint fragment. Filter hybridisation selection also resulted in the exclusion of a number of normal chromosome ends with telomeres containing variant repeats TGAGGG, TTGGGG and TCAGGG. Consequently, this additional telomere repeat array enrichment step was used to generate an ordered array library containing telomere-junction clones from individual AJ, with a suspected terminal deletion at 22q13.3. A candidate terminal deletion breakpoint clone was isolated from this library. The telomere-adjacent DNA of this clone mapped to chromosome 22 and showed significant sequence identity (96.2 %) to a BAC clone mapped to 22q13.3, the region to which the breakpoint was originally mapped by RFLP analysis (Nesslinger *et al.*, 1994). The sequence at the candidate breakpoint amplifies adjacent to an array of telomere repeats in patient AJ but not in 89 other individuals tested, including the patient's parents, indicating that it is a *de novo* telomere repeat array. A terminal location for this telomere repeat array has not been verified, but amplification of products of up to 2.3 kb strongly suggests that it is at a chromosome terminus. Ten nucleotides of unknown origin appear to have been inserted before the start of the telomere repeat array. The mechanism of chromosome healing in AJ is unknown, but the lack of both subterminal sequence at the breakpoint and variant repeats within the telomere repeat array strongly suggests that healing occurred by the *de novo* addition of telomere repeats to non-telomeric sequence, probably by telomerase. There are no features of note in the region around the breakpoint, apart from a high density of *Alu* repeat elements.

Introduction

The telomere-anchored PCR method was used successfully to isolate a terminal deletion breakpoint fragment from patient FB336R, but did not yield candidate breakpoints from six further patients (described in chapter 3). The telomere-anchored PCR method can be efficient, with sometimes over 50 % of clones positive for telomere repeat arrays, however this efficiency was not always reproducible, probably due to variations in the size enrichment and PCR steps which are beyond control. The main problem with this approach was the variable, and sometimes low, number of telomere positive clones within a library (12 % in ROBA), representing mainly normal chromosome ends previously isolated. An additional telomere-selection step is likely to increase the efficiency of cloning telomere repeat arrays, particularly single chromosome ends. In addition, many normal chromosome ends have been shown to amplify well in a telomere-anchored PCR reaction, such as the PGSEGB subterminal repeat, which is found at high copy number in all libraries screened. It would therefore be advantageous if the telomere-selection step reduced the number of clones containing normal chromosome ends by selecting telomeres with only TTAGGG repeats, including breakpoints healed by *de novo* telomere addition.

In order to achieve this, a filter hybridisation strategy was developed to select for TTAGGG repeat arrays, based on a method used to isolate human simple repeat loci (Armour *et al.*, 1994). This method involved filter hybridisation to tandemly repeated targets to rapidly enrich for tri- or tetranucleotide tandem arrays within the human genome, and the isolation of a large number of repeat-containing clones. Hybridisation screening of un-enriched genomic libraries had previously been used to isolate simple tandem repeat loci (Edwards *et al.*, 1991; Kalaitsidaki *et al.*, 1992; Weissenbach *et al.*, 1992) but was very inefficient, and required the screening of large numbers of clones to identify each positively hybridising clone in small insert libraries (Armour *et al.*, 1994). After enrichment by filter hybridisation, fragments were cloned to produce a library in which about 30% of clones contained trimeric or tetrameric tandem repeats (Armour *et al.*, 1994). This is less efficient than the telomere-anchored PCR based strategy, however, it was thought that in combination with the existing size fractionation and telomere-anchored PCR strategy, an additional filter hybridisation step would significantly increase enrichment for telomere repeat arrays.

This hybridisation selection procedure was initially tested using genomic DNA from patient FB336R, as the terminal deletion breakpoint at 7q32 had been isolated

by screening and sequencing an ordered array library generated using the telomere-anchored PCR strategy (N. J. Royle, unpublished data).

Aims of work described in this chapter

The aim of work discussed in this chapter was to develop a hybridisation selection strategy to use in combination with the size fractionation and telomere-anchored PCR strategy, to increase enrichment for telomere repeat arrays prior to cloning. This strategy would then be tested on DNA from a patient (FB336R) where the breakpoint had been isolated, and if successful utilised to isolate a terminal deletion breakpoint from a patient with a deletion of 22q13.3.

Results

Section 1. Filter hybridisation selection

5.1 Outline of the filter hybridisation selection step

An outline of the strategy is shown in figure 5.1. Briefly, the modified strategy involved the generation of small nylon filters bearing either synthetic arrays of combined variant tandem repeats TGAGGG, TCAGGG and TTGGGG, or TTAGGG tandem repeats only. Genomic DNA was digested with *Mbo*I, fragments larger than 5 kb size-selected by two rounds of preparative gel electrophoresis, and ligated to the Sau3AI-linker. These fragments were then amplified by 10 cycles of telomere-anchored PCR, and products hybridised to the filter bearing variant repeat types. The aim of the variant repeat filter was to remove telomere repeat arrays containing these variant repeat types, including many normal chromosome ends (Allshire *et al.*, 1989; Baird *et al.*, 1995). However, telomeres often contain other variant or degenerate repeats which would not be removed by this step. DNA not bound to the variant repeat filter was then recovered from the hybridisation buffer, denatured again, and hybridised to the second filter bearing TTAGGG telomere repeats. This aim of this filter was to have high selectivity for PCR products containing telomere repeats, in particular arrays with predominantly TTAGGG repeat types. DNA bound to this filter was then recovered, amplified by telomere-anchored PCR and cloned into pBluescriptII SK+ vector (Stratagene). Recombinant clones were picked into an ordered array library that was screened and sequenced as previously described. The modified strategy is described in detail in chapter 2 (Materials and Methods - Construction of telomere-anchored PCR libraries using hybridisation selection.).

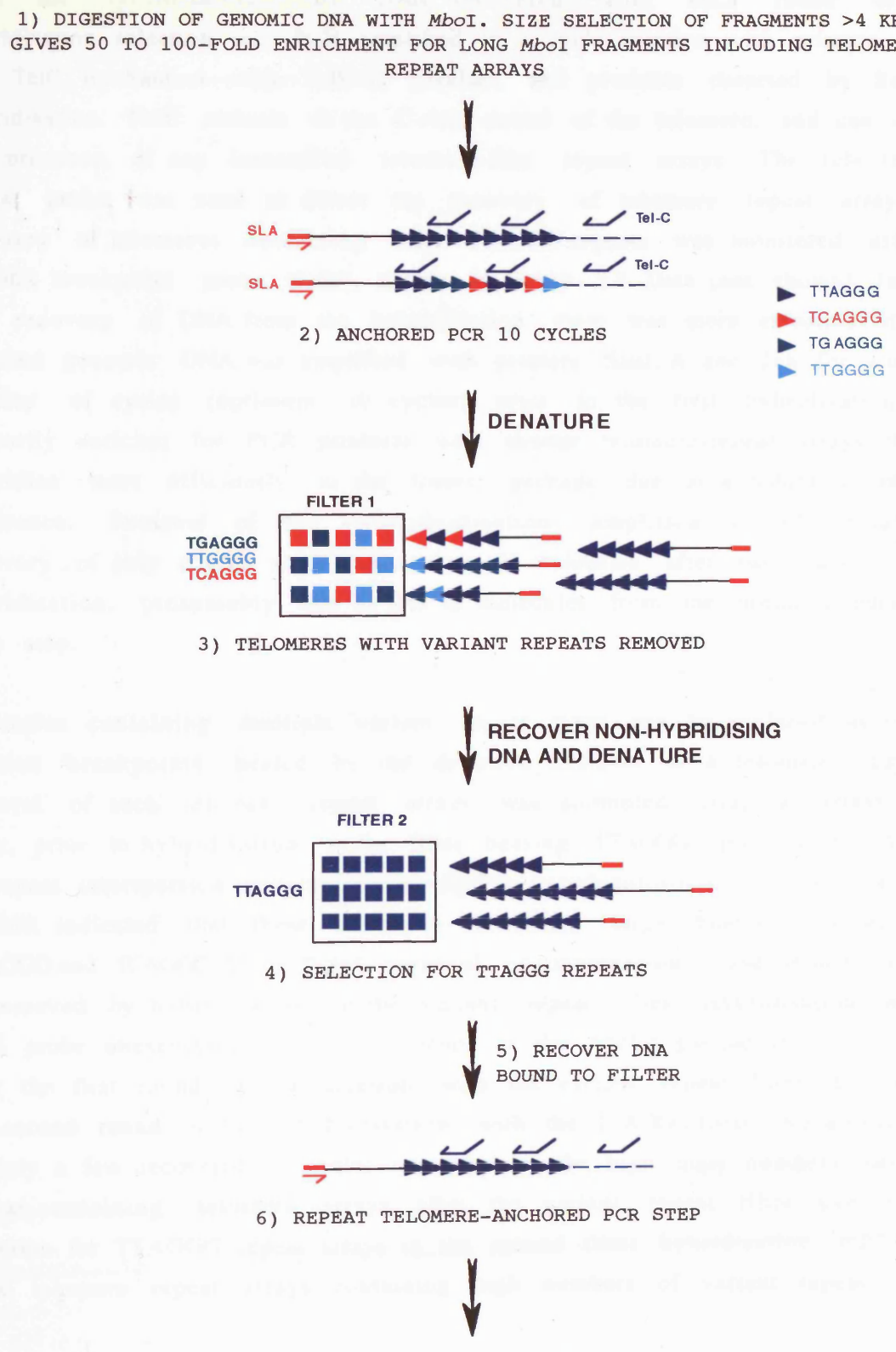
5.2 Developing and optimising hybridisation selection

The filter hybridisation selection step was optimised using genomic DNA from patient FB336R. This enabled assessment of the efficiency of the filter hybridisation step for selecting and enriching for telomere repeat arrays by comparison of data obtained from the original FB336R library (generated using telomere-anchored PCR), with data obtained using the modified strategy (including the hybridisation selection step). The nylon filters bearing TTAGGG repeats were hybridised with radioactively labeled Tel+ telomere repeat probe to ensure that synthetic repeats on the filter were available for binding DNA in the hybridisation mix. The strategy was tested at each stage to ensure that large amounts of DNA were not lost during each hybridisation, and that selection for

Figure 5.1 Schematic diagram of the filter hybridisation step used to enrich for telomere repeat arrays in ordered array libraries

1) High molecular weight genomic DNA is prepared by digestion with *Mbo*I. Size fractionation of products larger than 5 kb gives 50 to 100-fold enrichment for long *Mbo*I fragments including telomere repeat arrays.
2) Products from telomere-anchored PCR (10 cycles) with the SauL-A linker primer and the TelC telomere primer are recovered using a PCR purification spin column, alkaline denatured and neutralised.
3) Denatured PCR products are hybridised with the variant repeat filter to remove telomere repeat arrays containing variant repeats TGAGGG, TCAGGG and TCAGGG.
4) DNA not hybridising to the variant repeat filter is recovered, heat denatured and incubated with the TTAGGG filter.
5) PCR products bound to the TTAGGG filter are recovered by alkaline denaturation, neutralised and purified using spin columns.
6) Recovered PCR products are reamplified with linker primer SauL-A and telomere primer TelC, digested with *Eco*RI and *Kpn*I and cloned into pBluescriptII SK+ vector. Clones are arrayed in 96-well microtitre plates and replicated onto nylon filters.

Figure 5.1 Schematic diagram of hybridisation selection



telomere repeat arrays from long *Mbo*I-digested fragments was occurring. QIAamp spin columns (Qiagen, UK) were used for efficient recovery of the DNA from the hybridisation mix. DNA recovered after each round of filter hybridisation selection was PCR amplified in a cycle titration with primers SauL-A and TelC, or SauL-A with TelB in parallel, and products detected by Southern hybridisation. TelB anneals to the C-rich strand of the telomere, and can indicate the presence of any interstitial telomere-like repeat arrays. The Tel+ telomere repeat probe was used to detect the recovery of telomere repeat arrays, and recovery of telomeres containing only TTAGGG repeats was monitored using the FB336R breakpoint probe (6a8f), shown in figure 5.2. Data (not shown) indicated that recovery of DNA from the hybridisation steps was more effective if *Mbo*I-digested genomic DNA was amplified with primers SauL-A and TelC for a minimal number of cycles (optimum 10 cycles) prior to the first hybridisation. This primarily enriches for PCR products with shorter telomere-repeat arrays that will hybridise more efficiently to the filters, perhaps due to a reduction of steric hindrance. Removal of this pre-hybridisation amplification step resulted in recovery of only a very small population of molecules after two rounds of filter hybridisation, presumably due to loss of molecules from the initial population at each step.

Telomeres containing multiple variant repeat types can be excluded as terminal deletion breakpoints healed by the *de novo* addition of a telomere. Therefore removal of such telomere repeat arrays was attempted using a variant repeat filter, prior to hybridisation to the filter bearing TTAGGG repeats only. TVR maps of repeat interspersion patterns at the Xp/Yp pseudoautosomal telomere in patient FB336R indicated that these telomeres contained large blocks of repeat types TGAGGG and TCAGGG (D. M. Baird. personal communication), and should therefore be removed by hybridisation to the variant repeat filter. Hybridisation with the Tsk8 probe unexpectedly detected presence of the Xp/Yp pseudoautosomal telomere after the first round of hybridisation with the variant repeat filter, but not after the second round of filter hybridisation with the TTAGGG filter. Re-amplification of only a few recovered molecules may explain the high copy numbers of variant repeat-containing telomere arrays after the variant repeat filter hybridisation. Selection for TTAGGG repeat arrays in the second filter hybridisation may exclude some telomere repeat arrays containing high numbers of variant repeats.

However, it was not possible by PCR and Southern hybridisation alone to quantify enrichment for the breakpoint in the recovered molecules compared to the initial population, or to quantify any reduction in the number of Xp/Yp chromosome ends.

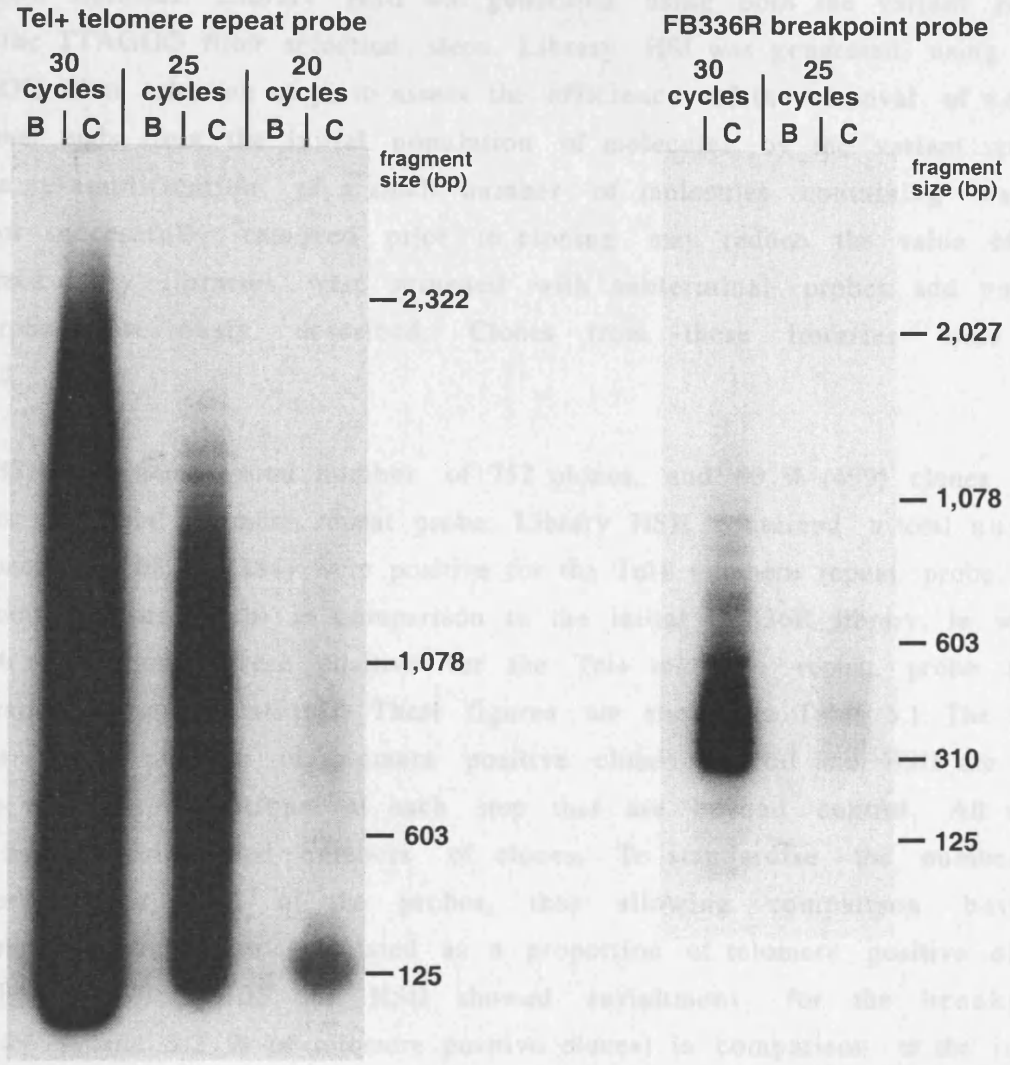


Figure 5.2 Optimisation of filter hybridisation selection using FB336R DNA

DNA recovered from two rounds of hybridisation selection using the variant repeat filter then the TTAGGG filter was PCR amplified using primers SauL-A and TelC (denoted by 'C' above each lane), or SauL-A and TelB (denoted by 'B') in a cycle titration reaction. Filter hybridisation selection resulted in efficient recovery of products containing telomere repeat arrays, detected by the Tel+ telomere repeat probe, and the FB336R breakpoint fragment, detected by the 6a8f breakpoint probe.

5.3 Generating libraries from patient FB336R using hybridisation selection

FB336R genomic DNA was prepared, as described, using hybridisation selection to generate two libraries. Library HSII was generated using both the variant repeat filter and the TTAGGG filter selection steps. Library HSI was generated using only the TTAGGG filter selection step, to assess the efficiency of the removal of normal chromosome ends from the initial population of molecules by the variant repeat filter. The re-amplification of a small number of molecules containing variant repeats not successfully removed prior to cloning may reduce the value of this step. Ordered array libraries were screened with subterminal probes and variant repeat probes previously described. Clones from these libraries were not sequenced.

Library HSI contained a total number of 752 clones, and 80 % (499) clones were positive for the Tel+ telomere repeat probe. Library HSII contained a total number of 564 clones, and 68 % (384) were positive for the Tel+ telomere repeat probe. This is a substantial improvement in comparison to the initial FB336R library, in which only 34 % (128) clones were positive for the Tel+ telomere repeat probe (N. J. Royle, personal communication). These figures are shown in Table 5.1. The small differences in the number of telomere positive clones in HSI and HSII are most likely due to slight variations at each step that are beyond control. All three libraries had varying total numbers of clones. To standardise the number of clones positive for each of the probes, thus allowing comparison between libraries, these figures were calculated as a proportion of telomere positive clones in each library. Both HIS and HSII showed enrichment for the breakpoint fragment (4.6 % and 5.2 % of telomere positive clones) in comparison to the initial FB336R library (1.6 % of telomere positive clones). Data shown in Table 5.1 indicates a small reduction in the numbers of clones containing the Xp/Yp telomere junction (Tsk8) and in the number of clones containing the Tsk46 and Tsk37 subterminal repeats in both HIS and HSII, compared to FB336R. This suggests that the variant repeat filter is effective in removing a number of Xp/Yp, Tsk37 and Tsk46 telomeres containing variant repeats TGAGGG, TCAGGG or TTGGGG. However, the HSI library would be expected to have similar numbers of these chromosome ends as the FB336R library due to the lack of variant repeat filter hybridisation step to remove these ends. These differences are most likely due to bias in the initial population of molecules, or variations at each step during library generation. The number of clones containing the PGSEGB subterminal repeat was not determined for the FB336R library. However, there were comparably high numbers of clones in HSI and HSII hybridising to this probe. The PGSEGB subterminal repeat is unlikely to be removed in significant numbers

Table 5.1 Probe hybridisation data from the FB336R library and hybridisation filter selection libraries HSI, HSII and AJ

The FB336R library was generated using the telomere anchored-PCR method. The HSI and HSII libraries were generated using the filter hybridisation selection method to enrich for telomeres within the population of molecules used to generate ordered array libraries. The TTAGGG telomere repeat filter only was used for selection in HSI. Both the TTAGGG and variant repeat filters were used for selection in HSII and AJ. Table 5.1 shows results of screening libraries with the Tel+ telomere repeat probe, subterminal probes and variant repeat probes. Blank boxes indicate that screening with a particular probe was not carried out. Positively hybridising clones for each probe are shown as the number of positive clones (no.), the percentage of positive clones (% of total) and the percentage of positive clones also positive for the Tel+ telomere repeat probe (% telomere positive). The latter figures are the most useful for comparison.

	FB336R Telomere-anchored PCR only			HSI Telomere-anchored PCR and hybridisation selection TTAGGG only			HSII Telomere-anchored PCR and hybridisation selection TTAGGG and variants			AJ Telomere-anchored PCR and hybridisation selection TTAGGG and variants		
Total size	376			752			564			752		
	no.	**% of total**	**% telomere positive**	**no.**	**% of total**	**% telomere positive**	**no.**	**% of total**	**% telomere positive**	**no.**	**% of total**	**% telomere positive**
Tel+	128	34		499	80		384	68		383	51	
breakpoint	2	0.5	1.6	23	3	4.6	20	3.5	5.2	12	1.6	3
Tsk37	28	7.5	21.9	70	9.3	14	35	6.2	9	38	5	10
Tsk46	6	1.6	4.7	2	0.3	0.4	7	1.2	1.8	46	6	12
Tsk8	5	1.3	4	5	0.7	1	1	0.2	0.3	0	0	0
E-F				32	4	6.4	15	2.7	3.9	10	1.3	2.6
CB0001-24c				6	0.8	1.2	1	0.2	0.3			
PGSEGB				198	26	39.7	163	28.9	42	59	7.8	15.4
TelG	22	5.8	17.2	278	37	56	84	15	22	64	8.5	26.7
TelJ	31	8.2	24.2	209	28	42	57	10	15	100	13.3	26
TelK	10	2.7	7.8	151	20	30	45	8	12	12	1.6	3

by the variant repeat filter, as previously sequenced telomeres adjacent to this repeat contained mainly other variant and degenerate repeats rather than TGAGGG, TCAGGG or TTGGGG at the proximal ends. However, longer telomere repeat arrays adjacent to PGSEGB may contain these variant repeat types, or TTAGGG repeats only. FB336R was not screened with the 7q/12q telomere-junction probe (E-F), however, HSII contained lower numbers of E-F positive clones than HIS, suggesting removal of these telomeres in the variant filter hybridisation step. Sequence from the proximal ends of the 12q and 7q telomeres from FB336R was obtained from telomere-anchored PCR products (J. Coleman, 1998). These regions contained mainly other variant repeats rather than TGAGGG, TTGGGG or TCAGGG types. However, the distal repeats are likely to contain variants, as telomere mapping at the 7q and 12q telomeres in other individuals showed that most of these telomeres contained some TCAGGG and TGAGGG repeat types. HSII showed a large reduction (more than half) in the number of clones containing variant repeats TGAGGG (TelG), TTGGGG (TelJ) and TCAGGG (TelK) in comparison to HSI (shown in table 5.1), confirming that the variant filter step is effective in removing a number of chromosome ends containing these variant repeats. However, HSII had similar numbers of variant repeat positive clones in comparison to FB336R which is unexpected due to the additional variant repeat filter hybridisation step. This is again probably due to slight variations at each step during library generation.

In summary, the hybridisation selection strategy is effective in enriching for telomere repeats, in conjunction with size fractionation and telomere-anchored PCR, particularly for telomere arrays containing only TTAGGG repeats. The general trends suggest a reduction in the number of clones containing normal chromosome ends using hybridisation selection, particularly when both the variant and TTAGGG filters were used.

Section 2. Using Filter Hybridisation to generate an ordered array library from patient AJ

5.4 Characterisation of AJ clones

Following successful enrichment for clones containing telomere repeat arrays, and the breakpoint fragment, in patient FB336R, filter hybridisation selection was used in conjunction with size fractionation of *Mbo*I-digested genomic DNA and telomere-anchored PCR to generate an ordered array library from patient AJ. A previously generated library from AJ generated by telomere-anchored PCR was discarded due to the presence of only one telomere-positive clone (described in

chapter 4). AJ has a phenotype of general developmental delay, hypotonia, a severe delay in expressive speech and mild facial dysmorphia. High-resolution cytogenetic analysis revealed a deletion of 22q13.3 (Nesslinger *et al.*, 1994). Replicate nylon filters were screened with the Tel+ telomere repeat probe, subtelomeric and variant repeat probes. Results of the screening are shown in table 5.1. The library contained a total number of 752 clones. 383 (51 %) clones were positive for telomere repeats. This figure was lower than expected given the success with the FB336R, but still efficient, particularly in comparison to the previous AJ library.

After screening with subterminal and variant repeat probes, 145 telomere positive clones remained to be sequenced. All clones contained the predicted sequence from the *Kpn*I to *Sau*3A sites of the linker unless otherwise stated. A total of 51 clones were sequenced, and are summarised in table 5.2.

5.5 Analysis of candidate breakpoint clone AJ-111b

Three identical clones AJ-111b, AJ-21g and AJ-29g, had 170 bp of telomere-adjacent sequence without homology to known subterminal regions. All three clones had an array of twenty-six TTAGGG repeats. The consensus sequence (called AJ-111b) from these identical clones is shown in figure 5.3. The clone structure met the criteria used for selecting candidate breakpoint clones, that is, telomere-adjacent sequence without homology to subterminal regions previously identified, and an array of TTAGGG repeats without variants. Consequently, AJ-111b was investigated as a candidate breakpoint region. The 170 bp of telomere-adjacent sequence from clone AJ-111b shows 96.2 % sequence identity over 160 bp to a region of BAC clone bk268H5 (EMBL/Genbank accession no. AL008718). The sequence alignment is shown in figure 5.4. It is possible that the poly(A) tail (bases 47 to 62) is polymorphic in the population, explaining the differences in this region between AJ-111b and bk268H5. The sequence identity with the genomic sequence from clone bk268H5 does not extend the entire length of the telomere-adjacent DNA, but terminates 10 bp proximal to the beginning of the telomere repeat array. The start of the AJ-111b telomere-adjacent sequence (base pairs 1 to 62) shows significant sequence identity to the *Alu*-Sx subfamily repeat consensus sequence (Jurka and Milosavljevic, 1991), shown in figure 5.4. At the time of writing, sequencing of BAC clone bk268H5 was in progress (classified as unfinished sequence) and may be subject to further change before completion. The sequence used for all analysis described in this chapter was obtained from the most recent update available (12th March 1999).

Clone name	No. of clones in group	Start site	Telomere-adjacent sequence	Sequence indentity FASTA/BLAST database search	Telomere repeats present	Variant repeats present	No. of repeat interspersion patterns in group	Previously identified?
GROUP 1	46	*Sau3A*	< 20 bp	_	yes	TGAGGG, TCAGGG, TTGGGG other variants	_	_
AJ-111b	3	*Sau3A*	170 bp	96.2 % BAC clone bk268H5 (AL008718) over 160 bp	26	no	1	_
AJ-16a		*Sau3A*	98 bp	87 % MLT1 repeat	8	one TTAGAG	_	_
AJ-28c		*Sau3A*	46 bp	96 % subterminal repeat (U53226)	yes	other variants	_	CB ROBAIIA

Table 5.2 Summary of clones sequenced from the AJ library

Clone names are shown in the first column. Clones are grouped according to similarity either in the telomere-adjacent DNA or telomere repeat array, if adjacent DNA is absent. The number of clones in each group, and the start site, Sau3A (GATC) or KpnI (GGTACC), are shown in columns two and three. The length of telomere-adjacent sequence in each group of clones is shown in basepairs. Results of sequence homology searches using BLAST and FASTA search tools at the EMBL/Genbank database are shown in the fifth column. Presence or absence of telomere repeats, and TGAGGG, TTGGGG, TCAGGG or other variant repeats is shown in columns six and seven. The number of interspersion patterns within the telomere repeat arrays of each group of clones is indicated in column eight. The far right-hand column indicates whether a clone has been previously isolated and from which ordered array library.

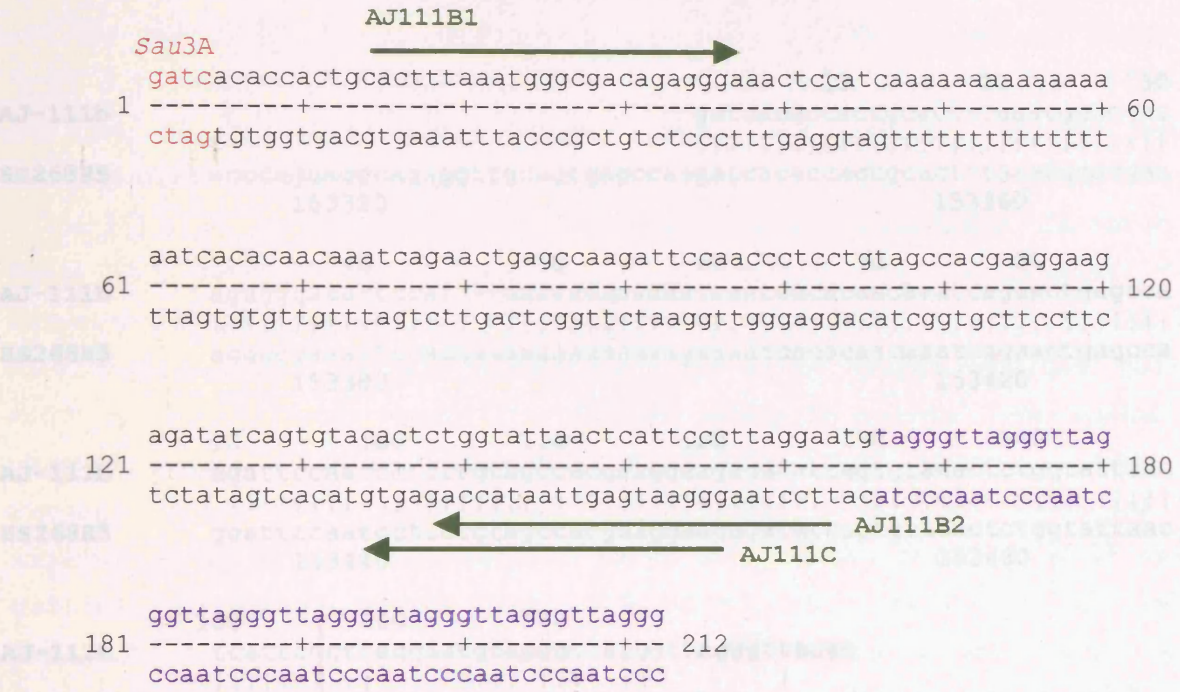


Figure 5.3 Consensus sequence of the putative AJ-111b breakpoint clone

Clone AJ-111b has 170 bp of telomere-adjacent DNA and an array of twenty-six telomere TTAGGG repeats, without variant repeats. The *Sau*3A site at the start of the clone is indicated in red, and the start of the telomere repeat array is shown in blue. Primers AJ111B1, AJ111B2 and AJ111C are indicated by green arrows. AJ111B1 and AJ111B2 amplify a 148 bp product.

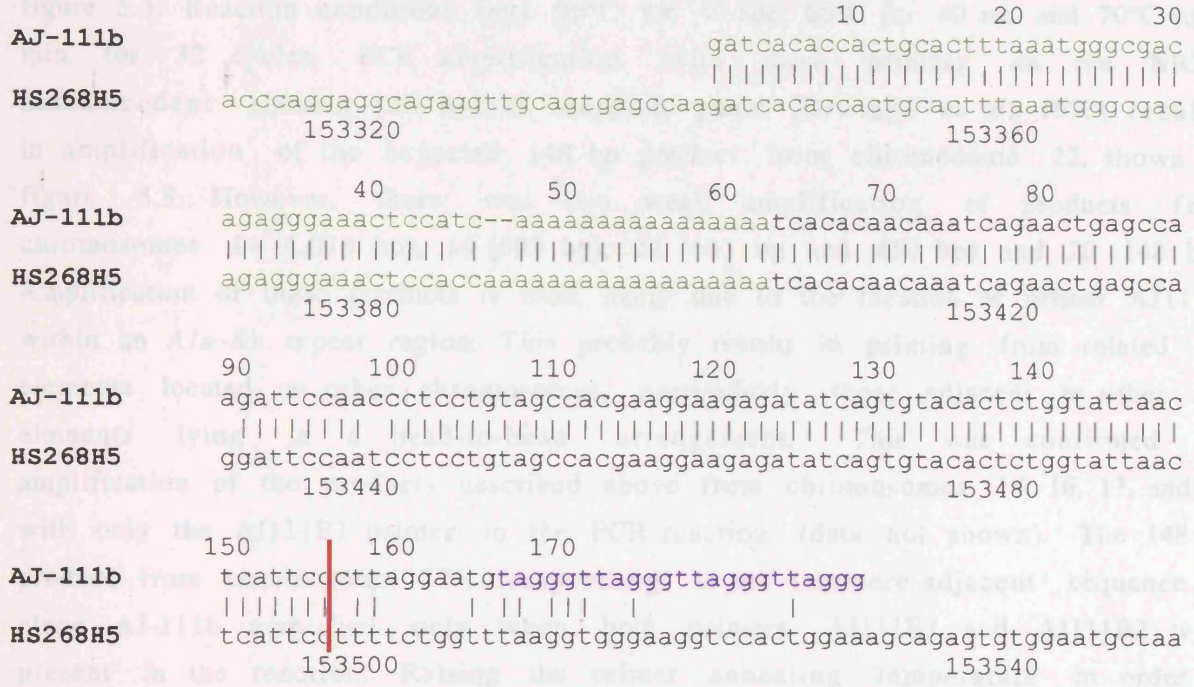


Figure 5.4 Alignment of the telomere-adjacent sequence from clone AJ-111b with a region of BAC clone bk268H5

The telomere-adjacent sequence from clone AJ-111b shows 96.2 % identity in 160 bp overlap with a region of BAC clone bk268H5 (EMBL/Genbank accession number AL008718). Bases showing matches between the two sequences are represented by vertical lines. Gaps have been inserted in the AJ-111b sequence to maximise the alignment, and are represented by dashes (-). The beginning of the AJ-111b telomere repeat array is shown in blue. The region of homology to the Alu-Sx element is indicated in green. The red vertical bar represents the distal limit of homology between AJ-111b and bk268H5 before the telomere repeat array.

5.6 Determining the chromosomal location of AJ-111b

BAC clone bk268H5 (AL008718) has been mapped to chromosome 22. To confirm the chromosomal location of the AJ-111b telomere-adjacent sequence, primers AJ111B1 and AJ111B2 were designed to generate a 148bp fragment, shown in figure 5.3. Reaction conditions were 96°C for 40 sec, 60°C for 40 sec and 70°C for 1 min for 32 cycles. PCR amplification with these primers on the NIGMS human/rodent somatic cell hybrid mapping panel (Drwinga *et al.*, 1993) resulted in amplification of the expected 148 bp product from chromosome 22, shown in figure 5.5. However, there was also weak amplification of products from chromosomes 14 (1,078 bp), 16 (603 bp), 17 (603 bp and 450 bp) and 20 (148 bp). Amplification of these products is most likely due to the location of primer AJ111B1 within an *Alu*-Sx repeat region. This probably results in priming from related *Alu* elements located on other chromosomes, particularly those adjacent to other *Alu* elements lying in a head-to-head arrangement. This was confirmed by amplification of the products described above from chromosomes 14, 16, 17, and 20 with only the AJ111B1 primer in the PCR reaction (data not shown). The 148 bp product from chromosome 22 corresponding to the telomere-adjacent sequence of clone AJ-111b amplified only when both primers AJ111B1 and AJ111B2 were present in the reaction. Raising the primer annealing temperature in order to increase the specificity of the PCR reaction was unsuccessful due to the high homology between related *Alu* elements, and eventually resulted in loss of amplification of the region on chromosome 22. The *Alu*-Sx region of AJ-111b also showed significant sequence identity to related *Alu* elements on chromosome 20, 14q31, 16q21 and others, consistent with amplification of these products from the monochromosome hybrid panel. These spurious products were gel purified and sequenced using the automated sequencing protocol as described in chapter 2 (Materials and Methods – Automated sequencing) with the AJ111B1 and AJ111B2 primers consecutively. However, this failed to generate readable sequence with either primer, except from the chromosome 22 amplicon. The poor sequence data appeared to be due to amplification from more than one target sequence, expected due to the repetitive nature of *Alu* repeats. The amplicon from chromosome 22 confirmed that the sequence obtained from clone AJ-111b containing the putative breakpoint region is also present in genomic DNA, and is not an artifact of cloning.

To obtain amplification of the candidate breakpoint region from chromosome 22 only, further primers were designed outside the *Alu* repeat regions to amplify a 850 bp product. Primer AJ111C was located in the cloned AJ-111b sequence, orientated away from the telomere repeat array. Primer AJ111D was located outside

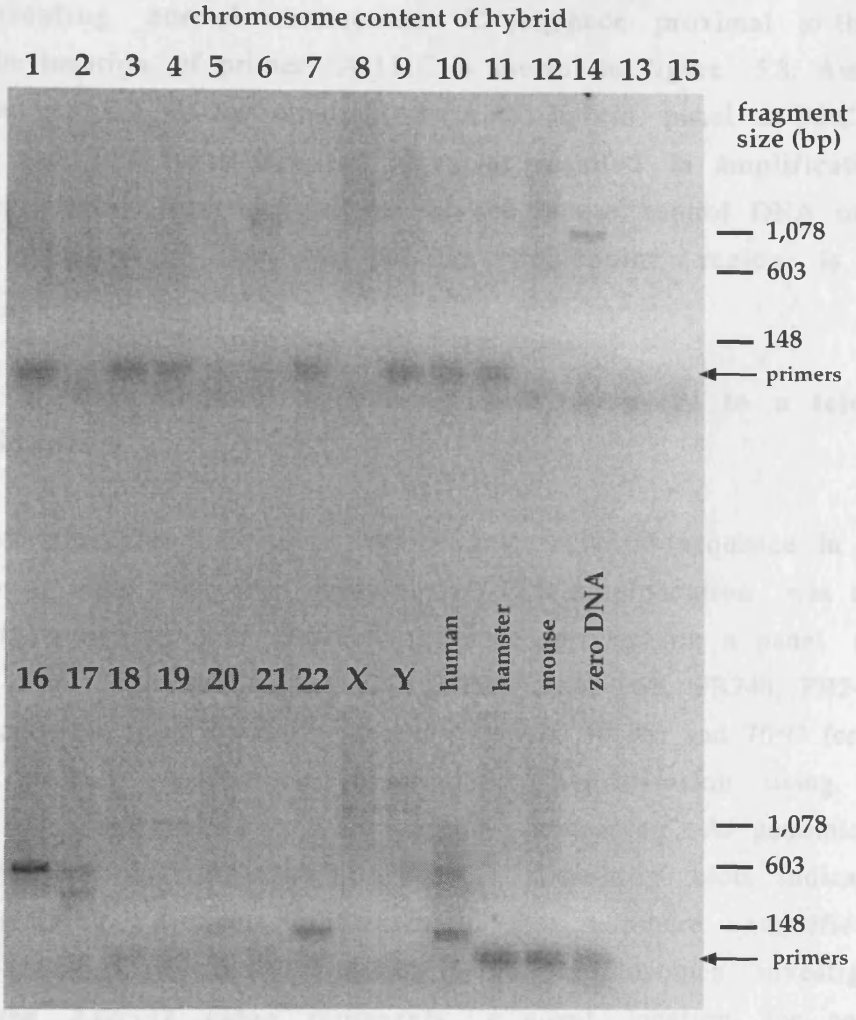


Figure 5.5 Mapping the telomere-adjacent sequence of AJ-111b on the monochromosome hybrid panel

PCR amplification was from the monochromosome hybrid panel using primers AJ111B1 and AJ111B2. Human, hamster, mouse and zero DNA controls were included. PCR products were separated on a 2.2 % HGT agarose gel. The expected 148 bp product amplified from chromosome 22, and the human control. There was also weak amplification of products from chromosomes 14 (1,078 bp), 16 (603 bp), 17 (603 bp and 450 bp), 19 (1,078 bp) and 20 (148 bp). Faint bands are also present in a number of other lanes. These products are thought to arise from primer AJ111B1, which lies within an *Alu* repeat. It is likely that primer AJ111B1 can amplify related *Alu* repeats from other regions of the genome.

the cloned region, and was designed from sequence obtained from BAC clone bk268H5, representing normal chromosome 22 sequence proximal to the putative breakpoint. The location of primer AJ111C is shown in figure 5.3. Amplification with AJ111C and AJ111D on the monochromosome hybrid panel at 96°C for 20 sec, 65°C for 40 sec and 70°C for 2 min for 30 cycles resulted in amplification of the expected 850 bp product from chromosome 22 and human control DNA only, shown in figure 5.6, confirming that the putative breakpoint region is unique to chromosome 22.

5.7 Location of the AJ-111b sequence with respect to a telomere in other individuals

In order to determine the location of the telomere-adjacent sequence in relation to a telomere repeat array in other individuals, PCR amplification was carried out using the AJ111B1 primer and the TelC telomere primer on a panel of 79 CEPH parents and 8 other unrelated individuals (P50, P53, P55, FB240, FB244, FB336R, KW). Cycling conditions were 96°C for 50 sec, 64°C for 40 sec and 70°C for 5 min for 20 cycles. PCR products were detected by Southern hybridisation using the AJ111B probe, which was generated in a PCR reaction containing AJ genomic DNA and primers AJ111B1 and AJ111B2 (148 bp product). Resulting blots indicated that a smear of hybridising products characteristic of telomere amplification was detected only in patient AJ, that is, in only 1/174 chromosomes investigated. This verifies that the AJ-111b clone represents a novel location for an array of telomere repeats. More significantly, the AJ-111b telomere-adjacent sequence does not amplify next to an array of telomere repeats in either parent, KJ and FJ, shown in figure 5.7. This confirms that the candidate breakpoint is a *de novo* telomere repeat array, and not an existing low level polymorphic telomere inherited from one of the parents. Given this data, it is highly likely that clone AJ-111b represents the terminal deletion breakpoint fragment. The discrete bands of approximately 800 bp detected in both parents, seen in figure 5.6, are most likely to be due to amplification of *Alu* elements at other genomic locations by primer AJ111B1. This was confirmed by the amplification of an 800 bp product from P50 genomic DNA with the AJ111B1 primer only. This product is not visible on the autoradiograph shown in figure 5.7 (20 PCR cycles) but is clearly visible after 25 cycles of PCR (data not shown). This also indicates that members of this *Alu* family do not normally lie immediately adjacent to telomeres in the genome. However, the putative AJ breakpoint sequence is clearly located adjacent to an array of telomere repeats in individual AJ only. This was verified by PCR amplification using TelC with primer AJ111D, which lies 855 bp proximal to the start of the telomere repeats, outside the AJ-111b cloned region. Individual AJ and parents KJ and FJ,

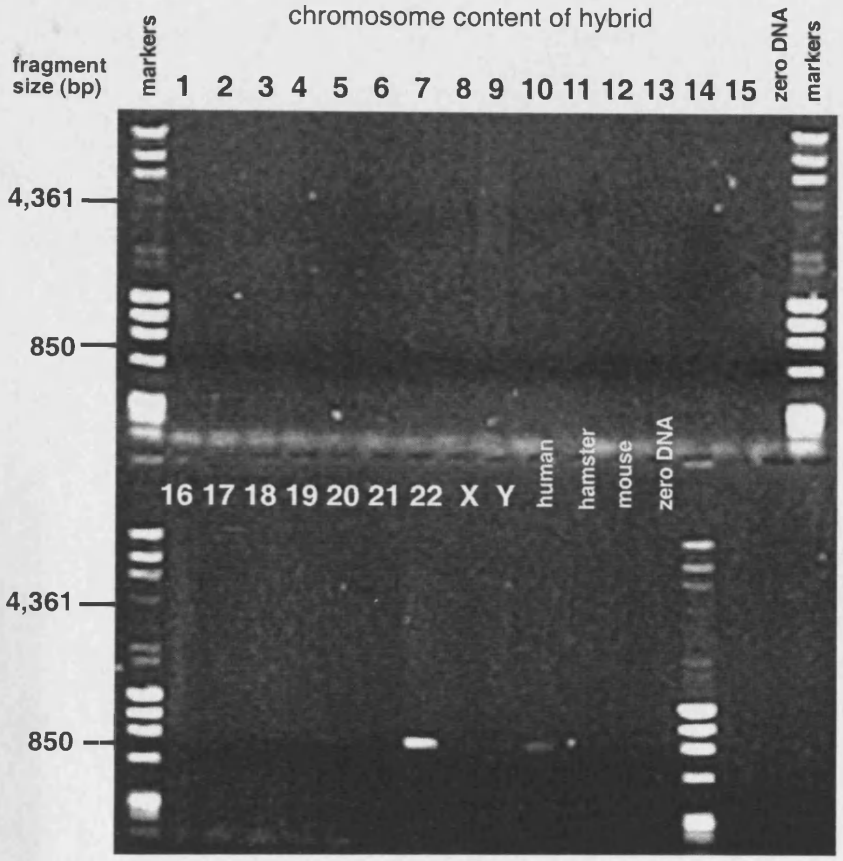


Figure 5.6 Mapping the AJ-111b telomere-adjacent sequence on the monochromosome hybrid panel using primers AJ111C and AJ111D

Primer AJ111C lies within the cloned region of the putative AJ breakpoint. Primer AJ111D lies outside this cloned region and was designed from sequence obtained from BAC clone bk268H5, representing normal chromosome 22 sequence proximal to the breakpoint. Human, hamster, mouse and zero DNA controls were included. The expected 850 bp product amplified from chromosome 22 and human control DNA only, confirming that this region is unique to chromosome 22.

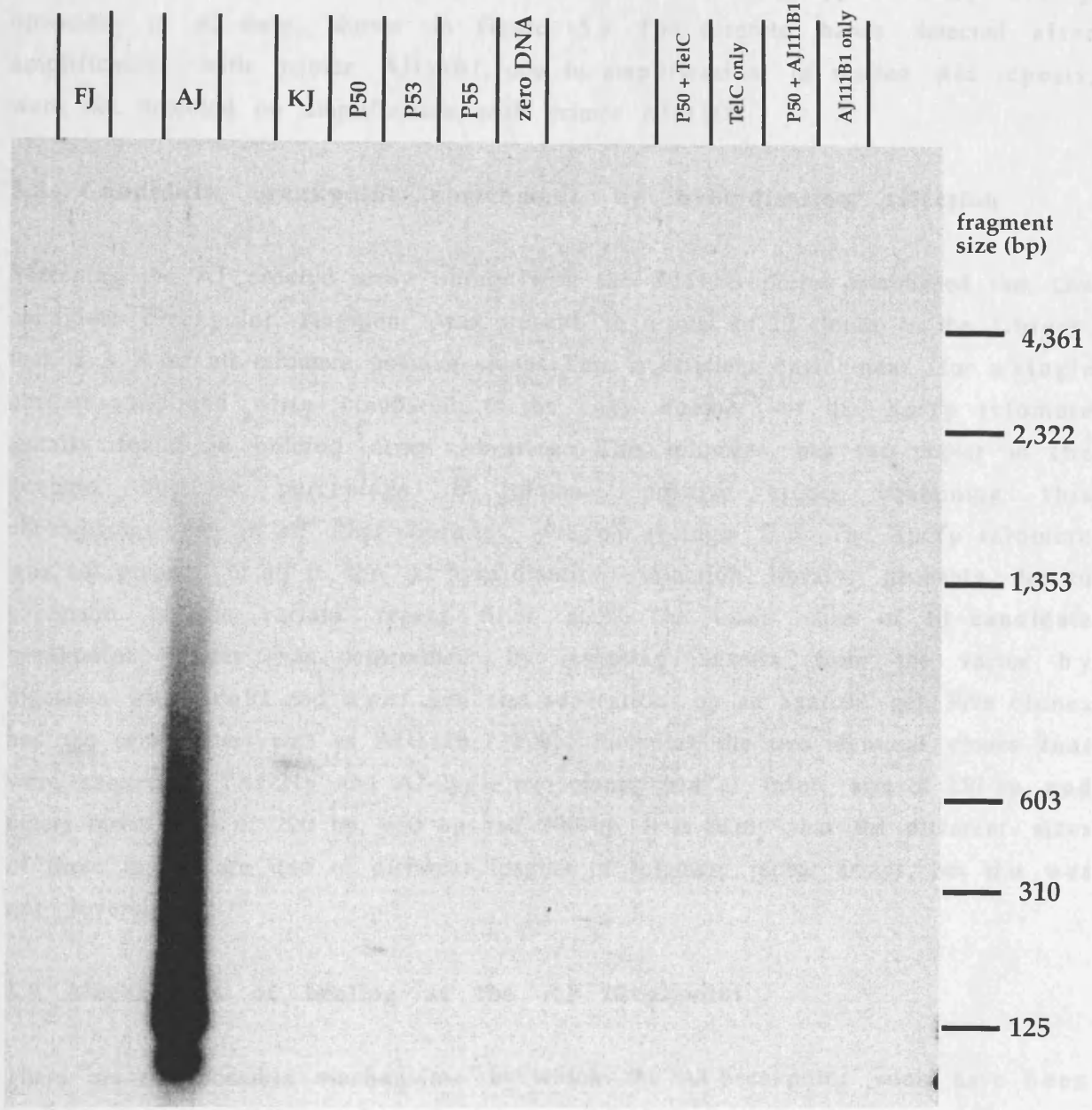


Figure 5.7 Amplification of the AJ-111b sequence adjacent to an array of telomere repeats in patient AJ

PCR amplification was carried out using primers AJ111B1 and TelC. Individuals used are indicated above each lane. Control reactions were also set up containing P50 DNA with only one primer, and single primer reactions without target DNA, indicated above. Southern hybridisation with the AJ111B probe detected the characteristic smear of hybridising products from a telomere repeat array in patient AJ only, but not in either of the parents, FJ and KJ, or three other unrelated individuals (P50, P53 and P55), indicating that this is a novel location for an array of telomere repeats. The presence of a faint, discrete product of approx. 800 bp in parents KJ and FJ is due to priming by AJ111B1 from Alu elements at other genomic locations.

plus P50, P53 and P55 were tested. Southern hybridisation with the AJ111B probe detected a smear of hybridising products of expected size (approximately 870 bp upwards) in AJ only, shown in figure 5.8. The discrete bands detected after amplification with primer AJ111B1, due to amplification of related *Alu* repeats, were not detected on amplification with primer AJ111D.

5.8 Candidate breakpoint enrichment by hybridisation selection

Screening the AJ ordered array library with the AJ111B probe established that the candidate breakpoint fragment was present in a total of 12 clones in the library, that is 3 % of all telomere positive clones. This is efficient enrichment for a single chromosome end when compared to the copy number of the Xp/Yp telomere usually found in ordered array libraries. This telomere has two copies in the genome, but the percentage of telomere positive clones containing this chromosome end in all other libraries was on average 2 %. The Xp/Yp telomere was not present at all in the AJ hybridisation selection library, probably due to exclusion by the variant repeat filter step. The insert size of 10 candidate breakpoint clones was determined by releasing inserts from the vector by digestion with *Eco*RI and *Kpn*I, and size separation on an agarose gel. Five clones had the same insert size as AJ-111b (330bp) including the two identical clones that were sequenced (AJ-21g and AJ-29g), two clones had an insert size of 250 bp, and others insert sizes of 220 bp, 460 bp and 700 bp. It is likely that the different sizes of these inserts are due to different lengths of telomere repeat arrays, but this was not investigated.

5.9 Mechanisms of healing at the AJ breakpoint

There are two possible mechanisms by which the AJ breakpoint could have been healed – *de novo* addition of a telomere, possibly by telomerase, or by capture of an existing telomere. *De novo* telomere addition would probably involve direct addition of repeats to the truncated chromosome onto non-telomeric sequence. If *de novo* telomere addition was mediated by telomerase, it is likely that this telomere repeat array would contain an array of TTAGGG repeats without variants. However, if healed by capture of an existing telomere, via a recombination event, the telomere repeat would most likely contain a number of variant or degenerate repeats. The AJ-111b clone contained twenty-six TTAGGG repeats without variants. However, it was possible that the genomic telomere repeat array outside this cloned fragment contained variant repeats. To look for presence of variants in the genomic telomere repeat array, TVR mapping was carried out on AJ genomic DNA with the AJ111B1 primer and one of the variant repeat primers TelG (TGAGGG),

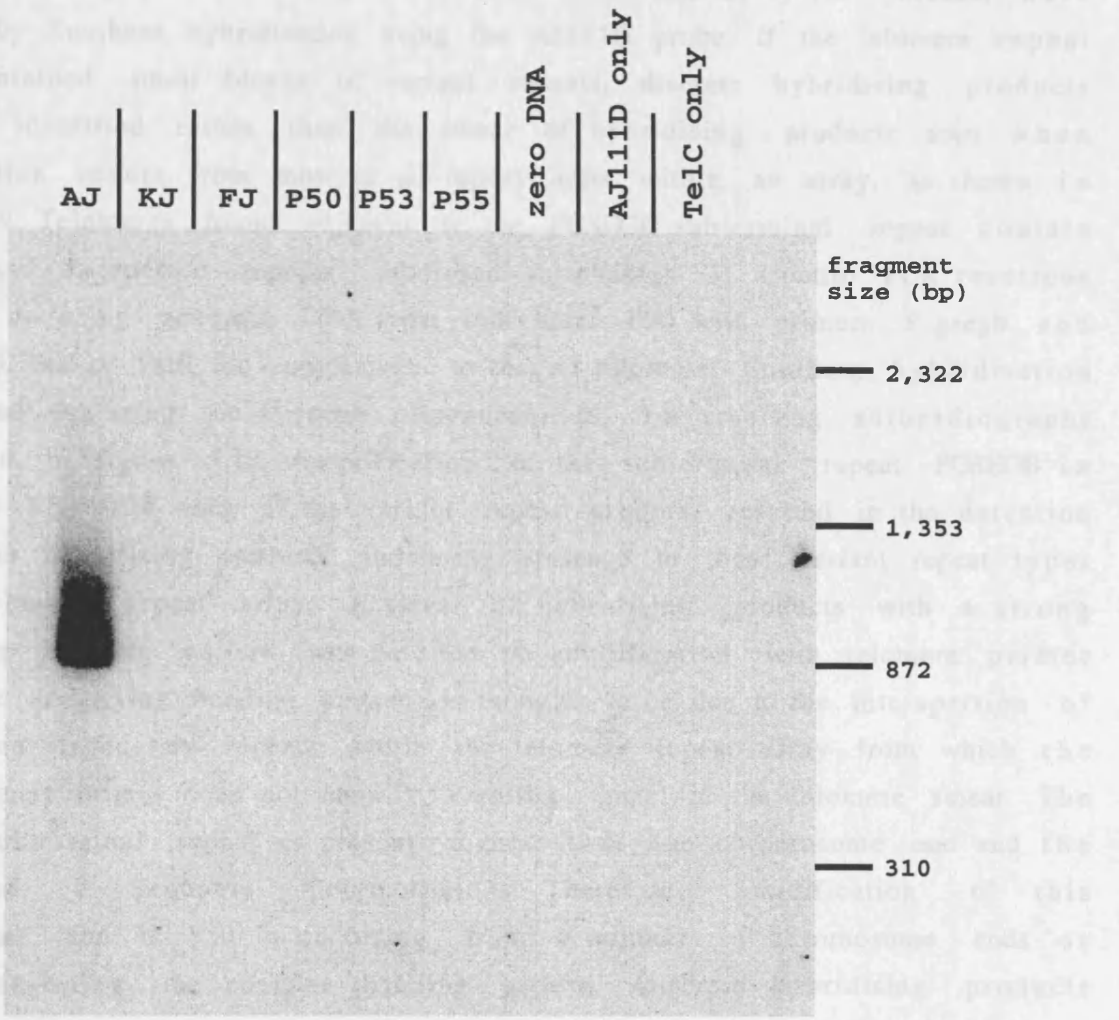


Figure 5.8 Amplification of the putative AJ breakpoint using primers AJ111D1 and TelC

Primer AJ111D was derived from the HS268H5 BAC clone, approximately 855 bp proximal to the breakpoint region in clone AJ-111b.Individual DNAs used are indicated above each lane. PCR amplification with primers AJ111D and TelC, followed by Southern hybridisation using probe AJ111B resulted in a smear of hybridising products ranging from the minimum expected product size of approximately 850 bp upwards in individual AJ only. This confirms that this region of chromosome 22q is adjacent to an array of telomere repeats in AJ but not in AJ's parents (KJ and FJ) or unrelated individuals. Control reactions containing zero DNA, or only one primer were included.

TelJ (TTGGGG) or TelK (TCAGGG) in separate reactions. Cycling conditions were 95°C for 20 sec, 64°C for 40 sec and 70°C for 5 min for 20 cycles. Products were detected by Southern hybridisation using the AJ111B probe. If the telomere repeat array contained small blocks of variant repeats, discrete hybridising products would be identified rather than the smear of hybridising products seen when amplification occurs from most or all repeat units within an array, as shown in figure 5.9. Telomeres found adjacent to the PGSEGB subterminal repeat contain variant and degenerate repeats (described in chapter 3). Control PCR reactions were set up using genomic DNA from individual P50 with primers P-gsegb and TelC, TelG, TelJ or TelK for comparison to the AJ telomere. Southern hybridisation was carried out using the P-gsegb oligonucleotide. The resulting autoradiographs are shown in figure 5.10. Amplification of the subterminal repeat PGSEGB in individual P50, with each of the variant repeat primers, resulted in the detection of discrete hybridising products, indicating presence of these variant repeat types in the telomere repeat array. A smear of hybridising products with a strong underlying banding pattern was detected on amplification with telomere primer TelC. The underlying banding pattern is thought to be due to the interspersion of variant and degenerate repeats within the telomere repeat array from which the TelC telomere primer does not amplify, creating 'gaps' in the telomere smear. The PGSEGB subterminal repeat is present at more than one chromosome end and the distribution is probably polymorphic. Therefore, amplification of this chromosome end in P50 is occurring from a number of chromosome ends or alleles, generating the complex banding pattern. Discrete hybridising products were not detected in AJ on amplification with variant repeat primers, and the characteristic telomere smear of hybridising products with TelC amplification had no detectable gaps, strongly suggesting that the putative breakpoint has an array of TTAGGG repeats only, without variant or degenerate repeats at the proximal end (up to 2.3 kb into the telomere). Therefore, the telomere at the putative AJ breakpoint is more likely to have been added *de novo*, possibly by the action of telomerase. The size of products amplified from this telomere, and the apparent lack of variant and degenerate repeats, strongly suggests that this repeat array has a terminal location. Interstitial telomere-like repeat arrays are generally less than 1 kb in size (Flint *et al.*, 1997), and variant or degenerate repeats are a common feature of these repeats (Wells *et al.*, 1990).

5.10 Location of the AJ breakpoint on the 22q map

The location of the deletion breakpoint in AJ was originally determined by high resolution cytogenetic analysis. The proximal limit of the deletion in individual AJ was determined using RFLP and dosage analysis with genetically mapped probes,

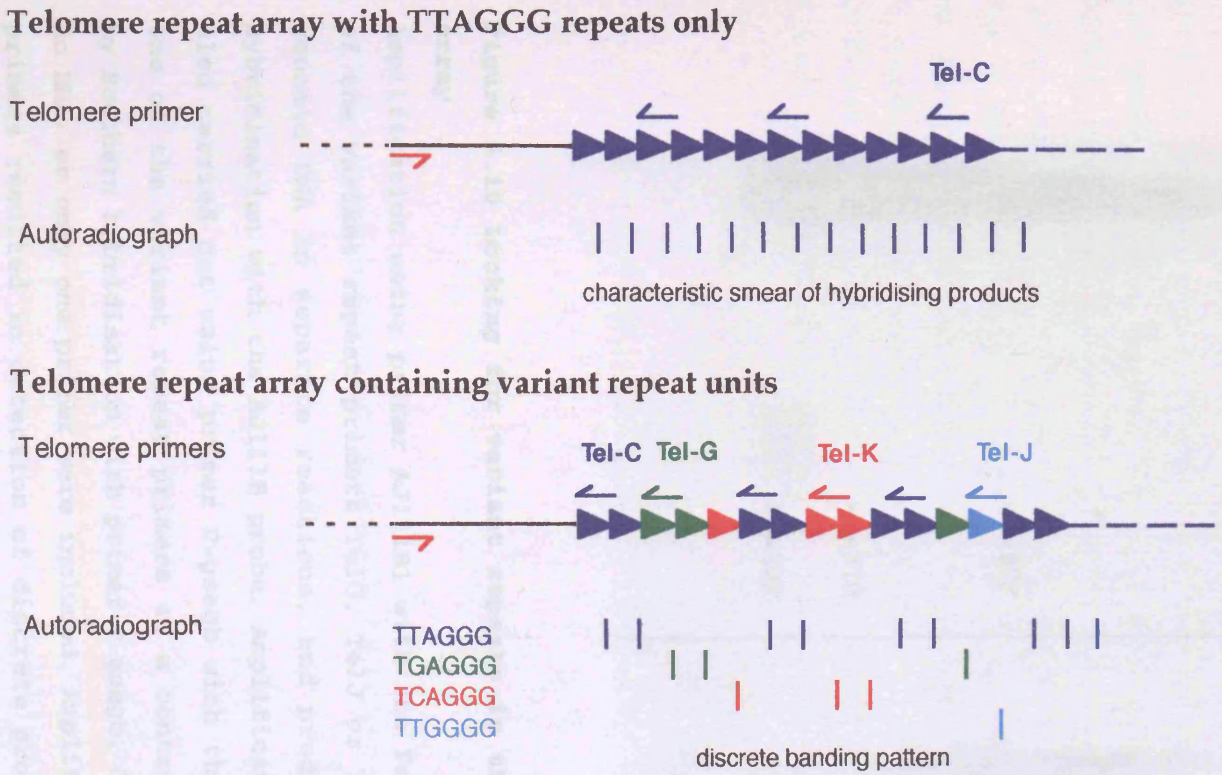


Figure 5.9 Schematic diagram showing identification of variant repeat units within a telomere repeat array by TVR amplification and Southern hybridisation

Amplification is carried out with a telomere-adjacent primer and the TelC telomere primer, or one of the variant repeat primers TelG (TGAGGG), TelJ (TTGGGG) and TelK (TCAGGG) in separate reactions. Products are detected by Southern hybridisation with a probe identifying the telomere-adjacent region of DNA. Amplification from variant repeats in the telomere repeat array results in discrete hybridising products, rather than a smear of hybridising products characteristic of amplification from most or all repeat units within an array.

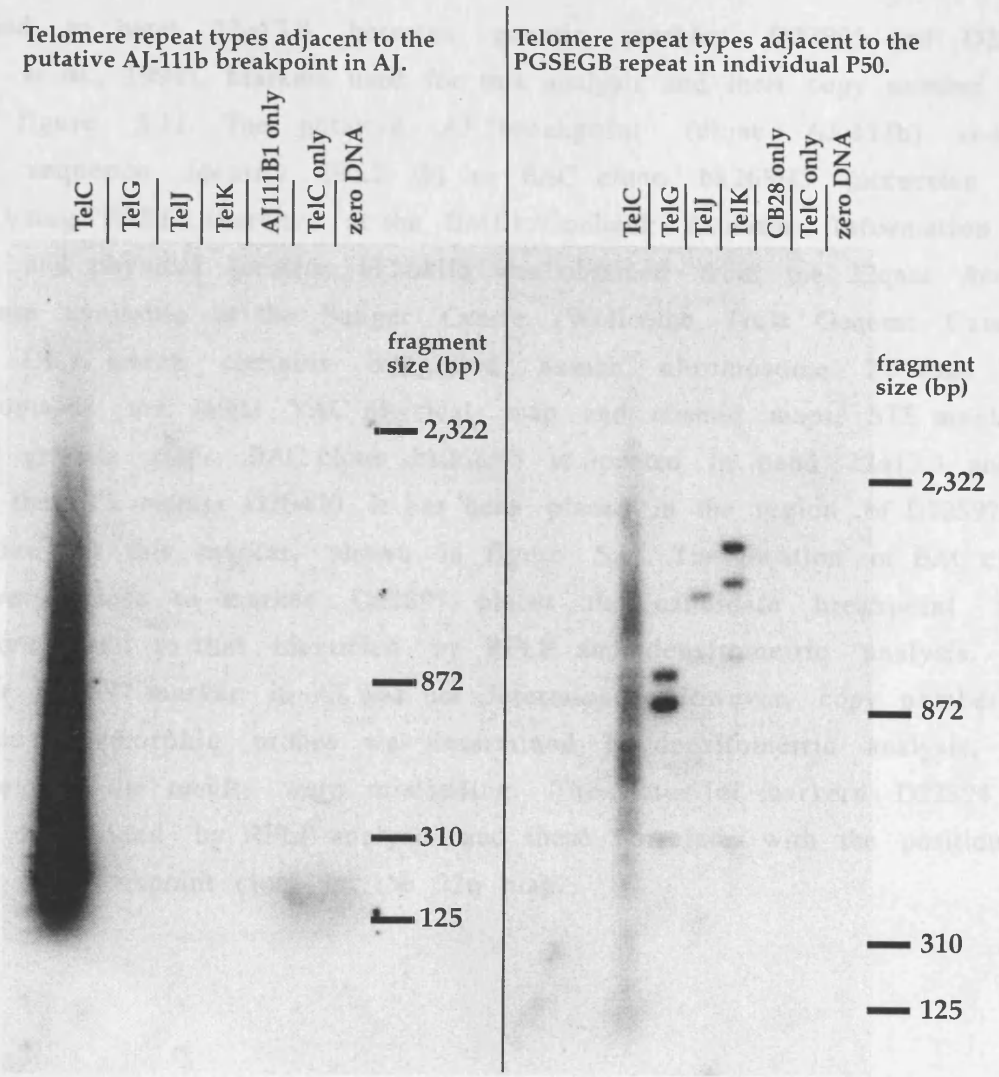


Figure 5.10 Looking for variant repeats in the AJ-111b telomere repeat array

Amplification using primer AJ111B1 with the TelC telomere primer, or one of the variant repeat primers TelG, TelJ or TelK was carried out on AJ genomic DNA in separate reactions, and products detected by Southern hybridisation with the AJ111B probe. Amplification of P50 genomic DNA was also carried out using primer P-gsegb with the TelC telomere primer, or one of the variant repeat primers as a control. Products were detected by Southern hybridisation with primer P-gsegb. Control reactions containing no DNA, or only one primer were included. Amplification with variant repeat primers resulted in detection of discrete products from variant repeats in the PGSEGB telomere repeat array. The putative AJ breakpoint telomere has a smear of hybridising products as expected with TelC amplification, but discrete products are not detected with variant repeat primers, suggesting that the telomere repeat array at the AJ breakpoint does not contain variant repeat types TGAGGG, TCAGGG or TTGGGG.

and localised to band 22q13.3, between genetic markers D22S95 and D22S64 (Nesslinger *et al.*, 1994). Markers used for this analysis and their copy number are shown in figure 5.11. The putative AJ breakpoint (clone AJ-111b) showed significant sequence identity (96.2 %) to BAC clone bk268H5 (accession no. AL008718) using FASTA searches at the EMBL/Genbank database. Information on the genetic and physical location bk268H5 was obtained from the 22qace AceDB-style database available at the Sanger Centre (Wellcome Trust Genome Campus, Cambridge, UK), which contains integrated human chromosome 22 data. This database contains the latest YAC physical map and cosmid maps, STS markers, probes and genetic maps. BAC clone bk268H5 is located in band 22q13.3 and is positive for the STS marker D20470. It has been placed in the region of D22S97 but is not positive for this marker, shown in figure 5.11. The location of BAC clone bk268H5, very close to marker D22S97, places the candidate breakpoint at a location more distal to that identified by RFLP and densitometric analysis. The status of the D22S97 marker in AJ was not determined. However, copy number of most of these polymorphic probes was determined by densitometric analysis, and it is possible that the results were misleading. The status of markers D22S94 and D22S95 was determined by RFLP analysis and these correlate with the position of the putative AJ breakpoint clone on the 22q map.

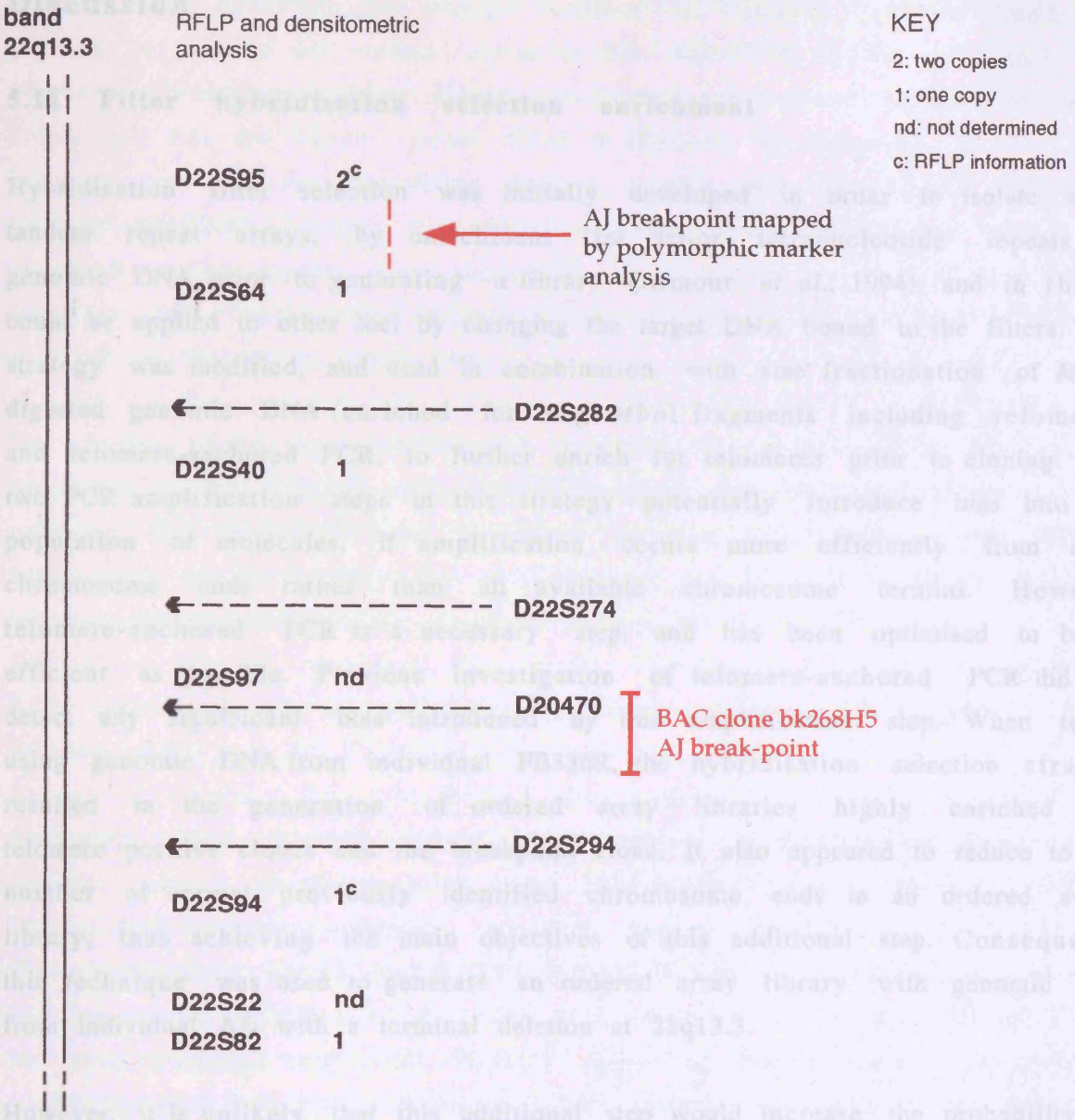


Figure 5.11 Mapping the AJ breakpoint by polymorphic marker analysis and sequence identity on the chromosome 22 map

Genetic markers are shown in the correct order but distances are not to scale. The AJ breakpoint was mapped by densitometric analysis of polymorphic markers to a region between markers D22S95 and D22S64 (Nesslinger *et al.*, 1994). The AJ-111b clone containing the putative AJ breakpoint sequence showed significant homology to a region of BAC clone bk268H5. This clone is positive for STS marker D20470, and has been placed in the same region as polymorphic marker D22S97. This suggests that the breakpoint lies at a location more distal to that identified by polymorphic marker analysis. All information shown in this figure was obtained from the 22qace database at the Sanger Centre, Oxford, UK (12/03/1999).

Discussion

5.11 Filter hybridisation selection enrichment

Hybridisation filter selection was initially developed in order to isolate short tandem repeat arrays, by enrichment for tri-or tetranucleotide repeats in genomic DNA, prior to generating a library (Armour *et al.*, 1994), and in theory could be applied to other loci by changing the target DNA bound to the filters. The strategy was modified, and used in combination with size fractionation of *Mbo*I-digested genomic DNA (enriched for long *Mbo*I fragments including telomeres) and telomere-anchored PCR, to further enrich for telomeres prior to cloning. The two PCR amplification steps in this strategy potentially introduce bias into the population of molecules, if amplification occurs more efficiently from some chromosome ends rather than all available chromosome termini. However, telomere-anchored PCR is a necessary step, and has been optimised to be as efficient as possible. Previous investigation of telomere-anchored PCR did not detect any significant bias introduced by this amplification step. When tested using genomic DNA from individual FB336R, the hybridisation selection strategy resulted in the generation of ordered array libraries highly enriched for telomere positive clones and the breakpoint clone. It also appeared to reduce to the number of normal previously identified chromosome ends in an ordered array library, thus achieving the main objectives of this additional step. Consequently this technique was used to generate an ordered array library with genomic DNA from individual AJ, with a terminal deletion at 22q13.3.

However, it is unlikely that this additional step would increase the probability of the isolation of large breakpoint fragments such as NT (3.5 kb), as these fragments tend to show inefficient PCR amplification, and they may be refractory to cloning in *E. coli* plasmid vectors. The number of telomere positive clones in the AJ library was lower than expected (51 %) in comparison to the pilot HSI (80 %) and HSII (68 %) libraries from patient FB336R, but was still efficient. A previously generated AJ library was not screened as it contained only 0.1 % telomere positive clones, and therefore could not be used for comparison. The AJ library showed a general trend of reduction in the number of previously isolated chromosome ends such as Tsk37, Xp/Yp (Tsk8), 7q/12q (E-F) and PGSEGB in relation to the FB336R and HSII libraries (shown in Table 5.1). However, the AJ library had a high copy number of Tsk46 in comparison to HSI and HSII, although variations in copy number of these subterminal repeats between individuals is expected, due to their polymorphic nature. The number of clones containing variant repeats was comparable in AJ, FB336R and HSII. In previously characterised ordered array libraries (CB0001,

ROBA, CB and CB00054), the average number of telomere positive clones also positive for any of the variant repeat probes was 72 %. In the AJ library, the number of telomere positive clones containing any variant repeats was 37 %, indicating that the variant repeat filter is effective in removing at least some telomere repeat arrays containing variant repeats (see table 5.1). A number of telomere positive AJ clones were negative for all the variant repeat probes but contained variant repeats when sequenced. However, the majority of these arrays contained only single variant repeats which are very unlikely to bind efficiently, if at all, to the variant repeat filter.

The addition of the filter hybridisation selection step to the general strategy was successful in the isolation of a putative terminal deletion fragment in AJ. Data indicates that it would have also been successful in isolating the FB336R breakpoint, and it is likely that this strategy can be used in future to isolate terminal deletion breakpoints in other patients.

5.12 Isolation of a candidate terminal deletion breakpoint from patient AJ

Screening of the AJ ordered array library resulted in the isolation of a putative terminal deletion breakpoint from three independent clones (consensus AJ-111b). The AJ-111b telomere-adjacent sequence mapped to chromosome 22 on the monochromosome hybrid panel. This region is located adjacent to an array of telomere repeats in individual AJ, but not in either of AJ's parents (KJ and FJ), or 187 other unrelated individuals, strongly suggesting that it is a novel location for a telomere. The lack of amplification in both parents demonstrates that it is not a low level polymorphic location for a telomere. Sensitivity of the telomere repeat array to *Bal*31 exonuclease digestion would confirm a terminal location. The telomere-adjacent sequence shows significant (96.2 %) sequence identity to a region of BAC clone bk268H5, which has been mapped to 22q13.3 on the chromosome 22 integrated map. RFLP and densitometric analysis of polymorphic markers also initially mapped the AJ breakpoint to this region (Nesslinger *et al.*, 1994), however, there is a small discrepancy over the precise location of the breakpoint. BAC clone bk268H5 is located in the region of marker D22S97, placing the breakpoint at a location more distal to that originally mapped (see figure 5.11). The order of markers in this region has remained the same since the densitometric analysis was carried out. However, sequencing of BAC clone bk268H5 is at present unfinished, and it is possible, although unlikely, that the contig of clones in this region of the chromosome will change, placing the breakpoint in the correct position. It is perhaps more likely that densitometric

analysis of markers gave misleading results, and this needs to be re-evaluated in order to resolve the discrepancy between these data. The location of the AJ-111b sequence does correlate with RFLP data initially used to place the breakpoint (Nesslinger *et al.*, 1994).

5.13 Mechanisms of healing at the AJ breakpoint

Healing by *de novo* telomere addition

Characterisation of terminal deletion breakpoints in six patients with α-thalassaemia (Flint *et al.*, 1994; Lamb *et al.*, 1993) established that the truncated chromosomes had been healed by direct addition of telomere repeats to non-telomeric sequence, possibly by telomerase. The sequence at each of these breakpoints is shown in figure 5.12. Human telomerase has been shown to recognise the sequence from one of these breakpoints (TI) as a substrate for the addition of TTAGGG repeats *in vitro* (Morin, 1991). In five patients (BO, CMO, IdF, TAT, and TI), the 3-4 bp immediately proximal to the telomere repeats show complementarity to the RNA template of telomerase. It was not possible in these patients to determine the exact base at which the break occurred, because the 3-4 nucleotides of the normal sequence at the breakpoint were identical to, and in register with, the telomere repeat array. In the sixth patient (IC), there was no overlap between the breakpoint sequence and the telomere repeat array and there appeared to be an insertion of two 'orphan' nucleotides (GT) preceding the telomere repeat array. It was not clear how these base pairs were acquired. The proximal telomere repeat array at the IC breakpoint contained one variant repeat (TTTAGGG), and it was suggested that this chromosome may have been healed by capture of an existing telomere, with consequent transfer of these two 'orphan' base pairs. However, it is possible that the variant repeat was generated by telomerase via slippage during repeat addition. Analysis of 200 nucleotides of the normal sequence either side of the α–thalassaemia breakpoints identified the pentanucleotide $(G)_5$ within 80 bp distal to all breakpoints, but the strand on which this pentanucleotide was found was not stated. None of the breakpoint regions contained TTAGGG repeats (Flint *et al.*, 1994).

The putative AJ breakpoint also appears to have been healed by the direct addition of a new telomere onto chromosome 22 unique sequence, but the structure of the telomere-junction is complex. Comparison of this region with sequence from a normal copy of chromosome 22 (BAC clone bk268H5) indicates that 10 'orphan' nucleotides have been generated between the end of the normal sequence and the beginning of the telomere repeat array (see figure 5.4). Therefore, it is not

Figure 5.12 Comparison of the putative AJ breakpoint with breakpoints from a number of other patients

The name of the individual, and chromosomal location of each breakpoint is shown in each case. The sequence across the breakpoint is shown on the top line, and the corresponding sequence from the normal chromosome on the bottom line, in a 5' to 3' orientation. Nucleotides showing matches between the sequences are represented by vertical lines. Telomere repeats (ttaggg) at the breakpoints are shown in blue. The sequence at the putative AJ breakpoint appears to have ten 'orphan' nucleotides (shown in red) inserted between the end of the normal sequence and the *de novo* telomere repeat array. In individuals BO, CMO, IdF, TAT, TI, the precise nucleotide at which telomere repeats were added could not be determined as the 3-4 nucleotides of the normal sequence at the breakpoint (underlined) were identical to and in register with the *de novo* telomere repeat array (Flint *et al.*, 1994). Similarly, in FB336R, the two nucleotides (underlined) of normal sequence at the breakpoint were identical to the *de novo* telomere repeat array. In individual LC, two 'orphan' nucleotides (shown in red) had been inserted between the end of the normal sequence and the telomere repeat array (Flint *et al.*, 1994). In NT the precise nucleotide at which the telomere repeats were added is identified when the breakpoint sequence is aligned with the normal sequence.

Figure 5.12 Comparison of the putative AJ breakpoint with breakpoints from a number of other patients

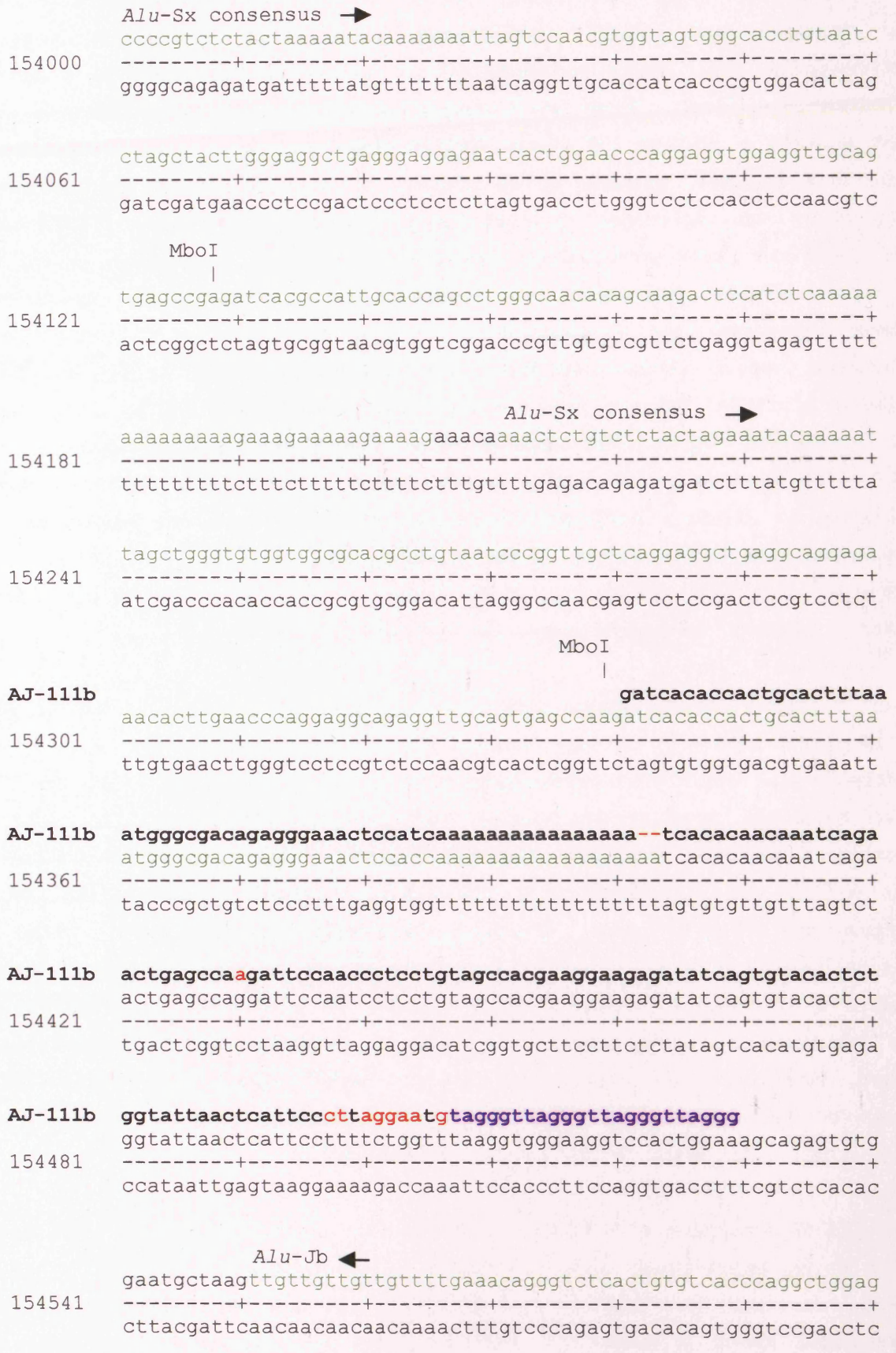
```
individual  deletion   5'                                     3'

AJ          22q13.3    tcattcccttaggaatgtagggttagggttagggtt
                       ||||||| ||     | || |||  |
                       tcattcctttctggtttaaggtgggaaggtccact

NT          22q13.3    agggggtggagaggggggtggtggagagggttagggtt
                       |||||||||||||||||||||||||| ||| | | |
                       agggggtggagaggggggtggtggagggggtggtggc

BO          16p13.3    gtgtgtacagagaagcaaggggttagggttagggtt
                       ||||||||||||||||||||||||       || |
                       gtgtgtacagagaagcaagggggtgctttctgccgct

CMO         16p13.3    agccgaggttctgtgttcagaggttagggttagggtt
                       ||||||||||||||||||||||||      |    |
                       agccgaggttctgtgttcagaggtcacctggaacaat

IdF         16p13.3    aagtggggctggggaacctccagggttagggtaggg
                       ||||||||||||||||||||||||   |||   ||
                       aagtggggctggggaacctccaggacgggggaggggc

TAT         16p13.3    ccccagggtccccttatatacagggttagggttaggg
                       ||||||||||||||||||||||||    ||     |
                       ccccagggtccccttatatacagggagctgggcccag

IC          16p13.3    ccaccgcgctcagccatggtggttagggttagggtta
                       ||||||||||||||||||   |      || |  |
                       ccaccgcgctcagccatgcccggctaattttgtatt

TI          16p13.3    gtgggtggggcgagcagagttagggttagggttaggg
                       |||||||||||||||||||||  | |      |||
                       gtgggtggggcgagcagagttgagctcccctccagga

FB336R      7q32       gaaatatggccacgagagggttagggttagggttaggg
                       ||||||||||||||||| ||     |    |  |
                       gaaatatggccacgagaaggaacctgggcttggcaata
```

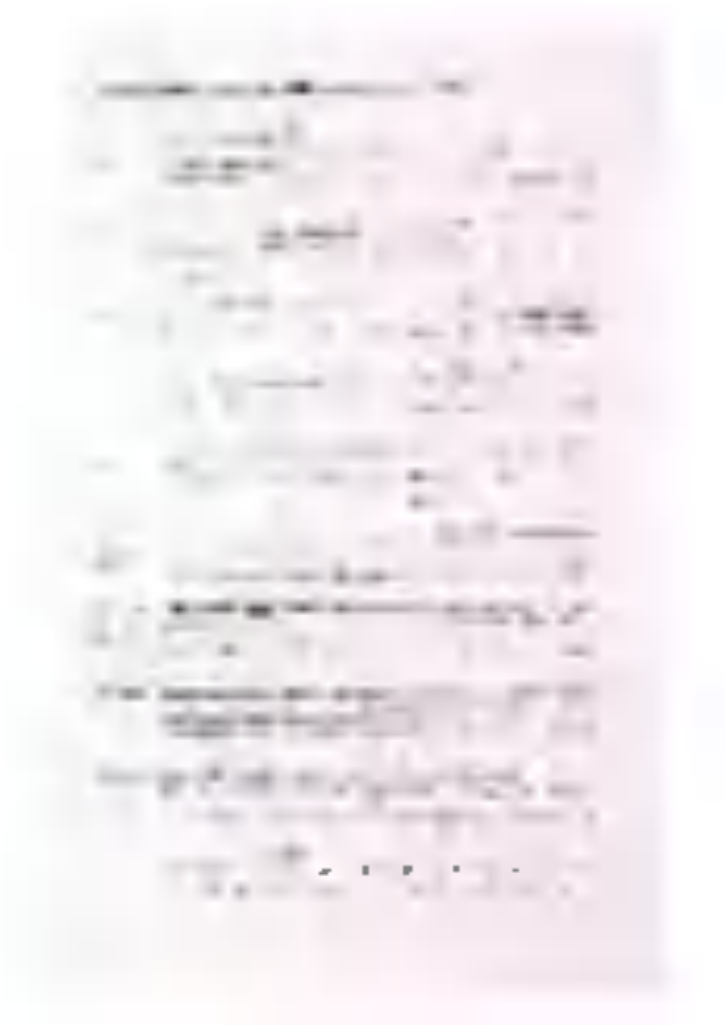
possible to predict at exactly which nucleotide the telomere repeats were added. It is not clear how the 10 'orphan' nucleotides at the AJ breakpoint may have arisen. It is possible that they were already present in AJ, and the differences between the truncated and normal chromosomes result from a common sequence polymorphism, which could be established by sequencing this region in a number of normal individuals. There are a number of other possibilities. The break may have occurred at position 159 of the AJ-111b clone, and these 'orphan' nucleotides were then acquired, possibly by the addition of incorrect repeats by telomerase. Nine of the ten nucleotides show similarity to degenerate repeats (TTAGGAATG), and could have been generated by telomerase 'stuttering', or slippage, at the start of repeat addition, generating new sequence at the break before repeats were added. Alternatively, the 10 nucleotides could have been added by a combination of telomerase repeat addition and the action of another enzyme at the breakpoint (Flint *et al.*, 1994). Normal sequence extending approximately 400 bp either side of the breakpoint region was obtained from BAC clone bk268H5 (accession no. AL008718). There is a high density of *Alu* repeats in this region (shown in figure 5.13), and an *Alu*-Sx extends into the cloned region of the putative breakpoint. The breakpoint lies within 100 bp of the 3' end of this *Alu*. A feature of note is the pentanucleotide $(G)_5$ 159 bp distal to the breakpoint. There were no TTAGGG repeats found in the 400 bp region around the AJ breakpoint.

Short sequence regions around other terminal deletion breakpoints that appeared to have been healed by *de novo* telomere addition have also been examined. The NT breakpoint lies within minisatellite locus MS607 at 22q13.3 (Wong *et al.*, 1997). The exact nucleotide at which the telomere repeats were added is identified when the breakpoint sequence is aligned with the sequence from a normal chromosome in this region, shown in figure 5.12. The sequence around the breakpoint did not contain any TTAGGG repeats. However, the breakpoint was located within the hexanucleotide $(G)_6$ on the strand corresponding to the G-rich telomere strand. Only a short stretch of normal sequence distal to the breakpoint was available, but the pentanucleotide $(G)_5$ occurs twice within 60 nucleotides on the G-rich telomere strand. Proximal to the breakpoint, $(G)_5$ also occurs three times within 200 bp (twice on the G-rich strand, once on the C-rich strand), and the hexanucleotide $(G)_6$ occurs twice within 80 bp (one on each strand).

The FB336R breakpoint at 7q32 was isolated by the telomere-anchored PCR method (N. J. Royle, unpublished data). Alignment of the breakpoint sequence with the normal sequence identified two nucleotides where telomere repeats might have been added, shown in figure 5.12. Analysis of normal sequence 200 bp either side of the breakpoint did not identify any TTAGGG repeats. The pentanucleotide $(G)_5$

Figure 5.13 Structure of BAC clone bk268H5 in the AJ breakpoint region





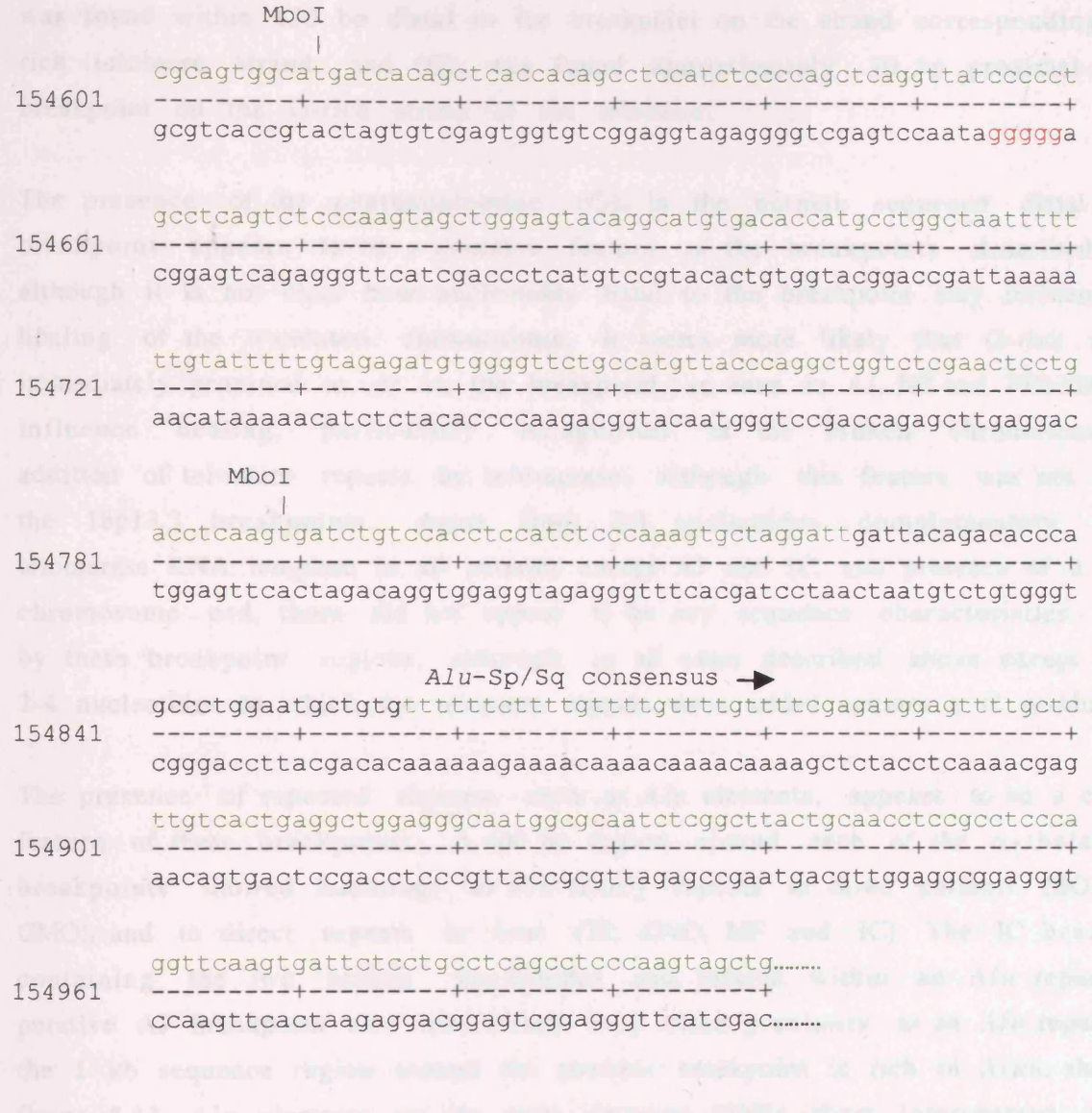


Figure 5.13 Structure of BAC clone bk268H5 in the AJ breakpoint region

The 1 kb region of BAC clone bk268H5 (representing 22q genomic sequence) around the breakpoint shows a high density of *Alu* repeats. *Alu* sequences are indicated in green, and the subfamily is shown above each region of homology. The black arrows indicate the orientation of the Alu repeats relative to the bk268H5 sequence. The alignment of the putative AJ breakpoint sequence with this region of bk268H5 is indicated in bold type. The first 62 bp of the AJ-111b sequence shows a match to an *Alu*-Sx repeat. Base mismatches are shown in red, and dashes (-) have been inserted to maximise alignment. The start of the AJ telomere repeat array is shown in blue. The pentanucleotide $(G)_5$ located 149 bp distal to the breakpoint is highlighted in red.

was found within 150 bp distal to the breakpoint on the strand corresponding to C-rich telomere strand, and $(G)_4$ was found approximately 70 bp proximal to the breakpoint on the G-rich strand of the telomere.

The presence of the pentanucleotide $(G)_5$ in the normal sequence distal to the breakpoint appears to be a common feature of the breakpoints described here, although it is not clear how nucleotides distal to the breakpoint may influence the healing of the truncated chromosome. It seems more likely that G-rich regions immediately proximal to, or at, the breakpoint, as seen in AJ, NT and FB336R, could influence healing, particularly recognition of the broken chromosome and addition of telomere repeats by telomerase, although this feature was not seen at the 16p13.3 breakpoints. Apart from 2-4 nucleotides complementary to the telomerase RNA template in all patients except AJ and IC, and presence of a broken chromosome end, there did not appear to be any sequence characteristics shared by these breakpoint regions, although in all cases described above except AJ, the 2-4 nucleotides to which the telomere repeats were added contain a G residue.

The presence of repeated regions, such as *Alu* elements, appears to be a common feature of these breakpoints. A 400 bp region around each of the α–thalassaemia breakpoints showed homology to *Alu*-family repeats in three patients (BO, IC and CMO), and to direct repeats in four (TI, CMO, IdF and IC). The IC breakpoint containing the two 'orphan' nucleotides was located within an *Alu* repeat. The putative AJ breakpoint also lies within very close proximity to an *Alu* repeat, and the 1 kb sequence region around the putative breakpoint is rich in *Alu*s, shown in figure 5.13. *Alu* elements are the most abundant SINEs (short interspersed nuclear elements) in the human genome, existing in copy numbers of up to 1 million per human haploid genome, representing 10 % of all human DNA. *Alu* repeats are around 300 bp long, and consist of two related units with a poly(A) tail of 20-30 bases flanked by short, direct repeats (reviewed Batzer *et al.*, 1993). *Alu* repeats can be divided into distinct subfamilies based on different bases in a number of diagnostic sequence positions (Jurka and Smith, 1988). These repeats appear to frequently be involved in rearrangements due to recombination, and have been implicated in illegitimate recombination events causing number of human genetic disorders. This will be discussed in detail in the following chapter. However, the terminal deletion breakpoints in the 16p13.3 α-globin do not appear to be consistently related to other structural or functional elements found in this region, such as genes, CpG-rich islands or DNAse I-hypersensitive sites (Vyas *et al.*, 1992).

Healing by telomere capture

Truncated chromosomes can also be healed by the capture of an existing telomere. Direct evidence for this mechanism as a cause of human disease has been found. A deletion of the HS-40 remote regulatory element of the α-globin genes (16p) in patient MB resulted in a phenotype of α-thalassaemia. The truncated chromosome appears to have been stabilised by capture of an existing telomere from a homologous chromosome, or by an interstitital subterminal deletion (Flint *et al.*, 1996). The break occurred within an *Alu* element of the Sx sub-family, only 105 kb from the 16p subtelomeric region, and may have been stabilised by unequal exchange between two *Alu* elements orientated in the same direction, resulting in transfer of an existing telomere to a novel location. Sequence acquired distal to this breakpoint comprises known and new families of polymorphic subtelomeric repeats. The precise registration of misaligned *Alu* elements implies that these sequences play a significant role in the location of breakpoints (Flint *et al.*, 1996). It is likely that in most telomere capture events, subterminal sequence will be transferred with the telomere, and the existing telomere is likely to contain variant repeats. Consequently, truncated chromosomes healed by telomere capture would probably be excluded by the screening strategy used to identify terminal deletion breakpoints in ordered array libraries.

A similar telomere capture event may explain the presence of the 10 'orphan' nucleotides at the AJ breakpoint. These nucleotides may represent a small region of subterminal sequence transferred along with the captured telomere. This 'orphan' sequence is much larger than the 2 bp seen at the 16p13.3 deletion in individual IC (Flint *et al.*, 1994), although recombination events at minisatellites in the male germline have been shown to include complex conversion events. These events can result in scrambling of repeat types at or near the site of insertion, resulting in repeats in the recipient allele that have no obvious origin in either parental allele (Jeffreys *et al.*, 1995). However, the telomere-capture event occurring via *Alu-Alu* recombination described above suggests that the transfer of relatively large blocks of sequence occurs along with telomere repeat array, such as subterminal repeat regions (Flint *et al.*, 1996). Subterminal sequence was not found at the putative AJ breakpoint, and telomere repeats have been added directly onto chromosome 22 unique sequence. Also, the putative AJ breakpoint telomere repeat array does not contain any variant repeats in the proximal repeat array, and appears to consists of entirely TTAGGG repeats when PCR amplified. This data strongly suggests that it is a *de novo* telomere and is very unlikely to have been acquired by capture of an existing telomere (or by replication from an existing telomere), which will almost certainly contain variant repeat types.

The presence of both the orphan nucleotides and the apparently *de novo* telomere repeat array may indicate healing by a more complex scenario. The break may have been initiated by a recombination event, occurring at the position where the AJ and normal sequences diverge, with further breakage 10 bp distal to this point, where *de novo* telomere repeats were then added, perhaps by the action of telomerase. Complex cryptic rearrangements leading to chromosome deletions have been identified. A deletion breakpoint from an individual with 18q- syndrome, and a phenotype of mental retardation and dysmorphic features, has been cloned and characterised (Katz *et al.*, 1999). The chromosome did not carry a simple terminal deletion as expected, but appeared to have been healed by a complex event. The sequence across the breakpoint identified the addition of a 35 bp 'filler' sequence from an unknown location, followed by a satellite III DNA-containing telomeric fragment of size 475-1000 kb. The filler sequence and the satellite III DNA were not found on the normal chromosome 18. The presence of the 35 bp 'filler' sequence is characteristic of non-homologous recombination events of the non-immune system class (Roth *et al.*, 1989). Satellite III is normally found on the short arms of acrocentric chromosomes, but usually more than 1000 kb from the telomere. It is likely that this rearranged chromosome arose via non-homologous recombination with satellite III sequences, and a secondary rearrangement that shortened the satellite III translocated fragment (Katz *et al.*, 1999).

5.14 Evidence that clone AJ-111b represents the terminal deletion breakpoint

The structure of the putative AJ breakpoint was derived from a cloned PCR product. The isolation of three independent clones with this identical structure makes it highly unlikely that the insert sequence structure was due to clone rearrangement. The consensus telomere-adjacent sequence was confirmed by sequencing this region from genomic DNA amplified from chromosome 22 in the monochromosome hybrid panel. Amplification of the AJ-111b sequence adjacent to an array of telomere repeats from genomic DNA occurs in patient AJ only, and not in 89 other individuals including parents KJ and FJ, confirming that this is, without doubt, a novel location for an array of telomere repeats. Sequencing across the putative breakpoint from AJ genomic DNA rather than a cloned PCR product would confirm the exact location of the telomere repeat array. Analysis of sequence from this region from a number of normal individuals, including the patient's parents, may resolve some of the discrepancies between the AJ-111b sequence and BAC clone bk268H5, in particular the 10 'orphan' nucleotides at the putative breakpoint.

All data described in this chapter strongly suggests that the AJ-111b clone represents the terminal deletion breakpoint. Location of this region to 22q13.3 on the 22q integrated map agrees with the initial localisation of the breakpoint by RFLP analysis (Nesslinger *et al.*, 1994), although there is slight discrepancy in the exact location of the breakpoint. The lack of subterminal sequence at the breakpoint, and the array of TTAGGG repeats only, supports *de novo* addition of a telomere repeat array, possibly by telomerase, as the most likely mechanism for healing of the breakpoint in AJ.

Chapter 6

ANALYSIS OF SEQUENCE AROUND TERMINAL DELETION BREAKPOINTS AT 7q32, 22q13.3 AND 16p13.3

Summary

Analysis of sequence 200 bp either side of nine terminal deletion breakpoints at 7q32, 22q13.3 and 16p13.3 did not detect any sequence features of note. The only sequences common to the region around these breakpoints appeared to be a high density of *Alu* repeats. In seven out of nine cases, the 2-4 nucleotides at the breakpoint showed complementarity to the RNA template of telomerase. This chapter describes the analysis and comparison of the 6 kb sequence around breakpoints at 7q32 (FB336R), 22q13.3 (AJ), and three breakpoints at 16p13.3 clustered within a 6 kb region (CMO, TAT and IdF; Flint *et al.*, 1994). All of these breakpoints are located in GC-rich R-bands. No common sequence features were identified apart from a high density of interspersed repeats, particularly *Alu* elements. A 60 kb region around the AJ (22q13.3) breakpoint was also investigated. Two genes were present in the region distal to the breakpoint, the 40s ribosomal protein S10 fragment and the uroplakin III gene, which encodes a urothelial cell marker. These genes had not been mapped to this region previously. This 60 kb region was dense in dispersed repeat elements, which may be involved in illegitimate recombination events, resulting in chromosome breakage. It is not known what factors might predispose towards chromosome healing, particularly by telomerase, but no obvious features were found within the 6 kb regions of all breakpoints studied that might be implicated in this process.

Introduction

A number of terminal deletion breakpoints from chromosomes 16p, 22q and 7q have been characterised, and appear to have been healed by the direct addition of telomere repeats, probably by telomerase (described in detail in chapter 5). Analysis of the normal sequence 200 bp either side of six breakpoints at 16p13.3 (Flint *et al.*, 1994), one at 7q32 (FB336R), and two at 22q13.3 (AJ, NT; Wong *et al.*, 1997) did not reveal any common sequence features that may be expected to facilitate chromosome healing by telomerase, such as TTAGGG repeats, but most lie within, or in close proximity, to an *Alu* repeat (Flint *et al.*, 1994; data in chapter 5).

The telomeric band 16p13.3 is GC-rich, and contains a high frequency of unmethylated CpG dinucleotides (Harris and Higgs, 1993). The terminal 165 kb of 16p, proximal to the subterminal region, is gene-rich, as predicted, and contains four widely expressed genes in addition to the four globin genes and three pseudogenes (Vickers *et al.*, 1993; Vyas *et al.*, 1992). The presence of pseudogenes (representing integration events or unequal exchange) may indicate chromosomal instability in this region. This region is also rich in repeat elements such as SINEs, and contains a least five hypervariable minisatellites (Higgs *et al.*, 1989). It is possible that the structure of this region predisposes towards chromosome rearrangements and large deletions, although the majority of α-thalassaemias are due to point mutations or small deletions and duplications.

Alu repeat elements appear to be frequently involved in rearrangements at a number of loci (Lehrman *et al.*, 1985; Nicholls *et al.*, 1987; Rogers, 1985; Vanin *et al.*, 1983). Some cases of familial hypercholesterolaemia are caused by defects in the low density lipoprotein (LDL) receptor gene, resulting from deletions or duplications due to recombination between *Alu* elements in this region (Lehrman *et al.*, 1985; 1987). Eight of twelve large interstitial deletions in the human α-globin gene region clustered to one region, and *Alu* repeats were frequently found at the breakpoints (Nicholls *et al.*, 1987), strongly suggesting that these repeats played a role in these recombination events. *Alu* elements may promote homologous or illegitimate recombination, resulting in translocations or deletions (Nicholls *et al.*, 1987), which are then healed by telomere addition. Direct repeats, palindromic repeats, and interspersed repeats are also commonly found at recombination junctions in α-globin deletions and generally in illegitimate recombination events in mammals (Nicholls *et al.*, 1987). *Alu* elements are common in terminal bands, and it is possible that these elements are frequently involved in rearrangements purely due to their density. However, direct evidence for the involvement of *Alu* elements in illegitimate recombination and deletion

events has been found a patient with a terminal deletion causing α-thalassaemia. The truncated chromosome appears to have been healed by telomere capture. The break occurred within an *Alu* element of the Sx sub-family, and was stabilised by unequal exchange between two *Alu* elements orientated in the same direction. The precise registration of misaligned *Alu* elements suggests a significant role in the location of the breakpoint (Flint *et al.*, 1996). However, it is not known whether repeated regions are directly involved in the healing of broken chromosomes by addition of a new telomere by telomerase.

The NT breakpoint lies within minisatellite locus MS607 at 22q13.3 (Wong *et al.*, 1997). Some minisatellites, particularly GC-rich loci such as MS607 are unstable (Armour *et al.*, 1990; Vergnaud *et al.*, 1991). 80 % of hypervariable GC-rich minisatellites are clustered in proterminal regions, also rich in dispersed repeats (NIH/CEPH, 1992). Minisatellite repeat transfer can result in complex allelic rearrangements in the germline (reviewed Buard and Jeffreys, 1997; Jeffreys *et al.*, 1994; Jeffreys and Neumann, 1997), and it is possible that minisatellite recombination events may lead to chromosomal rearrangements, resulting in breakage. However, only 866 bp around the NT breakpoint have been sequenced (EMBL/Genbank accession number X58044) due to the difficulties encountered in sequencing GC-rich regions (J. A. L. Armour, personal communication).

Only 200 bp of sequence either side of the terminal deletion breakpoints isolated to date has previously been examined. It is possible that analysis of sequence over a larger region around these breakpoints may result in identification of common sequence factors that may be involved in chromosome breakage and healing in these individuals, although it is not known what features might be involved.

Aims of work described in this chapter

Comparison of 6 kb of sequence around breakpoints from patients FB336R (7q32), AJ (22q13.3), TAT, CMO and IdF (16p13.3; Flint *et al.*, 1994) was carried out to identify any common sequence features present in these regions external to the 200 bp previously characterised. Common sequences may indicate whether healing by telomerase-mediated *de novo* telomere repeat addition has any specific long-range sequence requirements, or whether some sequences are preferred over others. This analysis may also lead to identification of genes in close proximity to the FB336R and AJ breakpoints, and if so whether their deletion, or re-positioning adjacent to a telomere contributes to the phenotype.

Results

6.1 Subcloning the FB336R breakpoint region

Cosmid clone cDIZ was obtained from human genomic DNA cosmid library HL1095m (Clontech, USA), containing inserts cloned into the *Bam*HI site of vector pWE15. The cosmid library was subjected to three rounds of screening with probe 6a8f, which detects the region adjacent to the FB336R breakpoint, unique to chromosome 7. Eight cosmids containing this region were identified, two of which contained identical inserts (cA2Z and cDIZ). All eight cosmid clones were mapped to the correct chromosomal band, 7q32, by FISH (N. J. Royle, unpublished data). Cosmid cDIZ was selected for analysis, and double stranded DNA was prepared as described in chapter 2 (Materials and Methods – DNA extraction II). The cDIZ insert was approximately 30 kb in size. Digestion of cDIZ with restriction enzyme *Eco*RI released the vector as an 8 kb fragment, and yielded insert fragments of approximately 20 kb, 5.7 kb, 4 kb and 0.6 kb (data not shown). Southern hybridisation with the 6a8f breakpoint probe located the FB336R breakpoint region within the 5.7 kb *Eco*RI fragment. The 5.7 kb fragment was then subcloned into the *Eco*RI site of pBluescriptII SK+ vector (Stratagene) as described in chapter 2 (Materials and Methods – Construction of telomere-anchored PCR libraries) and double stranded plasmid DNA was extracted for sequencing.

6.2 Sequencing of the FB336R breakpoint region

Sequencing of the subcloned insert containing the FB336R breakpoint region was carried out by automated cycle sequencing as described in chapter 2 (Materials & Methods - Automated DNA sequencing). Primers M13forward (21 mer) and M13reverse (21 mer) were used to sequence from each end of the vector into the subcloned fragment. Primers 6a8fa and 6a8fb, located in the breakpoint region, were used to sequence across the breakpoint in both directions. Sequence data obtained using these four primers was used to design further primers in order to 'walk' along the entire 5.7 kb subcloned fragment. Sequence was obtained with each primer from at least two independent isolates of the subcloned fragment, and a consensus derived from these pooled sequences, although only a few minor base differences were seen, and probably reflect variations in cycle sequencing reactions. However, it should be noted that sequence across most of the subcloned fragment was derived from either the forward or reverse strand only. Descriptions of primers M13forward, M13reverse, 6a8fa, 6a8fb, FOR1, REV1, 6AA1, 6AA2, 6AA3, 6AB1 and 6AB2 are available in chapter 2 (table 2.1), and their location within the 5.7 kb subcloned *Eco*RI fragment is shown in figure 6.1.

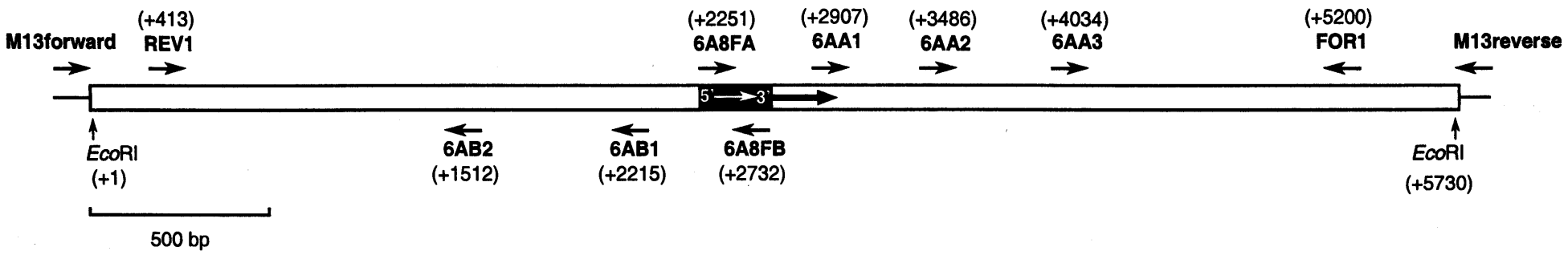


Figure 6.1 Location of primers used to sequence the 5.7 kb subcloned region of cosmid cDIZ containing the FB336R breakpoint region

FB336R breakpoint region

pBluescriptII SK+ vector

5.7 kb subcloned region from cosmid cDIZ

telomere repeats at breakpoint in FB336R

*Eco*RI cloning sites are indicated at each end of the subcloned region. Primers are indicated by black arrows and primer names are in bold type. The 5' *Eco*RI site is designated base position +1, and the 3' *Eco*RI site base +5730. The base positions of primer 5' ends are shown in brackets. The breakpoint region is shown as a black box, and lies in a 5' to 3' orientation. The location of the telomere repeat array in individual FB336R is indicated by a thick black arrow.

Sequence assembly was carried out as described in chapter 2 (Materials & Methods - Computer sequence analysis).

6.3 Sequence analysis of 5.7 kb around the breakpoint

The subcloned insert was 5730 bp in length, and the breakpoint region was orientated in a 5′ to 3′ direction with respect to the G-rich strand of the telomere repeat array found on the truncated chromosome. The orientation of the breakpoint sequence within genomic cosmid cDIZ was not determined, but could be confirmed by restriction mapping. The *Eco*RI site at the left end of the subcloned fragment was designated as base position +1 and the *Eco*RI site at the right end of the subcloned fragment was designated as base position +5730. The sequence found immediately adjacent to the telomere repeat array at the breakpoint (derived from clone 6a8f in the FB336R ordered array library) was located at base positions 2551 to 2850, shown in figure 6.1. The nucleotide composition of the region is A: 26.3 %; C: 22.8 %; G: 24.1 %, and T: 26.8 %, and the G-C content of the subcloned region is 47.3 %.

The NIX (Nucleotide Identify X) search tool at the HGMP Human Genome Resource Centre (Hinxton, Cambridge) was used to screen for putative coding sequences and exons within the subcloned region containing the FB336R breakpoint. The consensus did not show any significant sequence identity to nucleotide sequences or protein/peptide sequences present in the EMBL/Genbank databases, and no ESTs, STSs or CpG islands were identified. However, a 132 bp region (bases 3121-3253) gave an exon prediction in GRAIL/exons (probability 75 %) and GENEMARK (protein-coding exon probability 70 %). GENEFINDER and FGENE programs also gave gene predictions in this region. However, the prediction probabilities were not exceptionally high. Band 7q32 is well mapped, and as there were no protein or nucleotide sequence matches in the EMBL/Genbank databases, or any EST hits, it is unlikely that this sequence represents a true exon.

Repeat regions were detected using the RepeatMasker program (A. F. A. Smit and P. Green, unpublished data) via the NIX search tool. The RepeatMasker program is a repetitive element filter which screens DNA sequences against a library of human repetitive elements, masking all regions matching known repetitive elements including simple tandem repeats, low complexity regions, polypurine tracts and AT-rich regions. A summary of repeats found within the 6 kb around the FB336R breakpoint region is shown in table 6.1. Base positions 109-556, 826-895 and 1214-1274 showed a highly significant match to an L1 (LIME3) element in forward orientation. Positions 557-825 showed a highly significant match to an

	number of elements	length occupied	percentage of sequence
SINEs:	4	932 bp	16 %
ALUs	2	587 bp	10 %
MIRs	2	345 bp	6 %
LINEs:	1	579 bp	10 %
LINE1	1	579 bp	10 %
LINE2	0		
LTR elements:	1	547 bp	9 %
MaLRs	0		
Retrov.	0		
MER4	1	547 bp	9 %
DNA elements:	0		
MER1	0		
MER2	0		
Mariners	0		
Total interspersed repeats:		2058 bp	35 %
Small RNA:	0		
Satellites:	0		
Simple repeats:	0		
Low complexity:	0		

Table 6.1 Summary of repeats identified in the 5.7 kb sequence region around the FB336R terminal deletion breakpoint

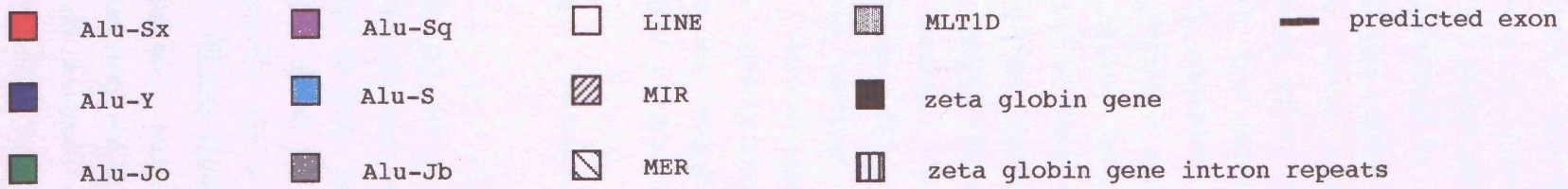
Data shown was obtained using RepeatMasker version 6/16/98db and ProcessRepeats version 06/16/98 via the NIX search tool at the HGMP Human Genome Mapping Centre. Most of the repeats fragmented by insertions or deletions are shown as one element.

Alu-Sx element in reverse orientation, and positions 896-1213 showed a highly significant match to an *Alu*-Jo element. Close analysis of this region indicated that it represents a single L1 element into which the *Alu* elements have inserted. Regions 4139-4327 and 5112-5267 showed a highly significant match to a MIR (mammalian-wide interspersed repeat) in reverse and forward orientation respectively. MIRs belong to the SINE family of repeats and are present in the human genome at a copy numbers of approximately 500,000 (Smit and Riggs, 1995). Positions 4485-5031 showed a highly significant match to a MER4 (medium reiterated frequency) sequence (Jurka, 1990) in a forward orientation. MER4 sequences are estimated to be present in the genome at a copy number of one to several thousand and are often associated with *Alu* elements (Jurka, 1989). There were no DNA transposable elements, satellites, simple repeats or low complexity repeats identified. However the program commonly fails to mask simple repetitive DNA elements that have different repeat patterns at each end. To look for such regions, the 6 kb sequence was compared to itself in both forward and reverse orientation by dot matrix analysis, using compare and dotplot programs in the Genetics Computer Group (GCG) sequence analysis software (Devereux *et al.*, 1984). This analysis compares nucleic acid sequence against another to visualise repeated or inverted-repeat structures. This analysis did not reveal any repeated regions other than expected similarity between *Alu* elements, and some very small regions of similarity of marginal significance. Sequence features around the breakpoint region are shown diagrammatically in figure 6.2. Further sequence from around the FB336R breakpoint region was not available, as this region has not yet been sequenced by the chromosome 7 mapping project.

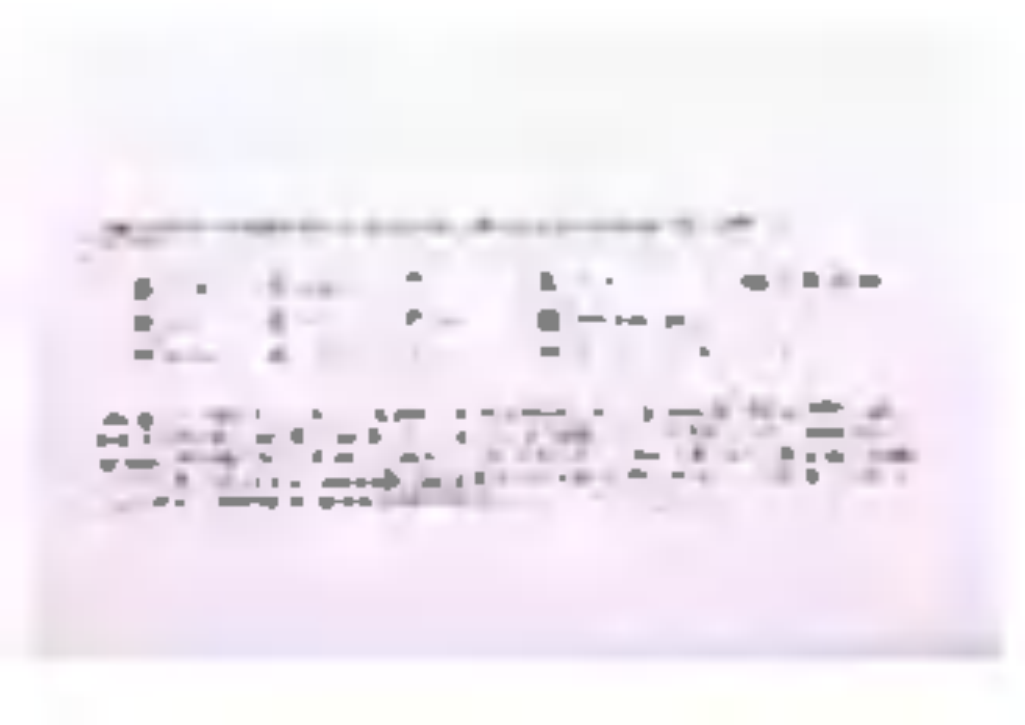
6.4 Location of the FB336R breakpoint region on the chromosome 7 physical map

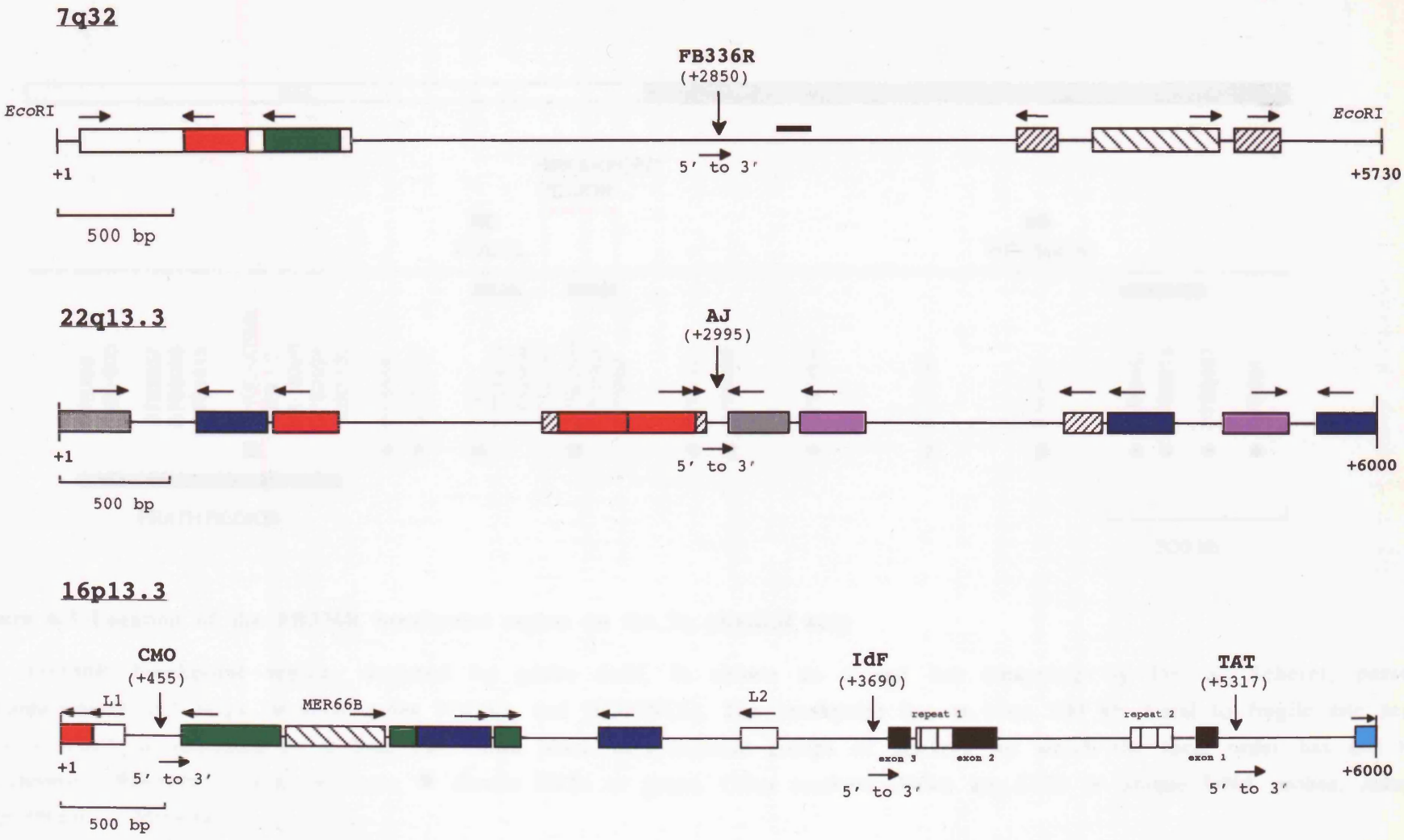
The sequence immediately adjacent to the FB336R breakpoint is detected using probe 6a8f (182 bp), and is unique to chromosome 7. The breakpoint region was mapped on the chromosome 7 physical map to 7q32-33 by hybridisation to a YAC clone contig of this region (mapping by Dr. S. Scherer, personal communication). The 7q physical map is available at the Chromosome 7 Project internet site (http://www.genet.sickkids.on.ca/chromosome7/). The breakpoint mapped to 7q32, distal to the PODXL gene (human podocalyxin-like membrane protein; EMBL/Genbank accession no. U97519), and proximal to the HIP/SNC6 cDNA clone in band 7q33, shown in figure 6.3. The common fragile site FRA7H lies within this region, centromeric to the PODXL gene. The FB336R breakpoint maps at least 500 kb distal to the fragile site. The FB336R breakpoint is located in the same region as STS markers D7S2386E, D7S2963 and D7S2362, however their local order has not yet

Figure 6.2 Sequence features in the 6 kb region around breakpoints in individuals FB336R, AJ , CMO, IdF and TAT



The FB336R breakpoint maps to 7q32. The putative AJ breakpoint maps to 22q13.3. Breakpoints CMO, IdF and TAT are clustered within a 6 kb region around the zeta globin gene at 16p13.3. The orientation of all breakpoints is shown with respect to the *de novo* telomere repeat array, and the base position at which the breakpoint is located within each 6 kb region is indicated. The orientation of repeat regions is indicated by arrows above each repeat. Sequence analysis was carried out using the NIX search tool at HGMP Human Genome Mapping Centre, Cambridge.





Chapter 6. Figure 6.2

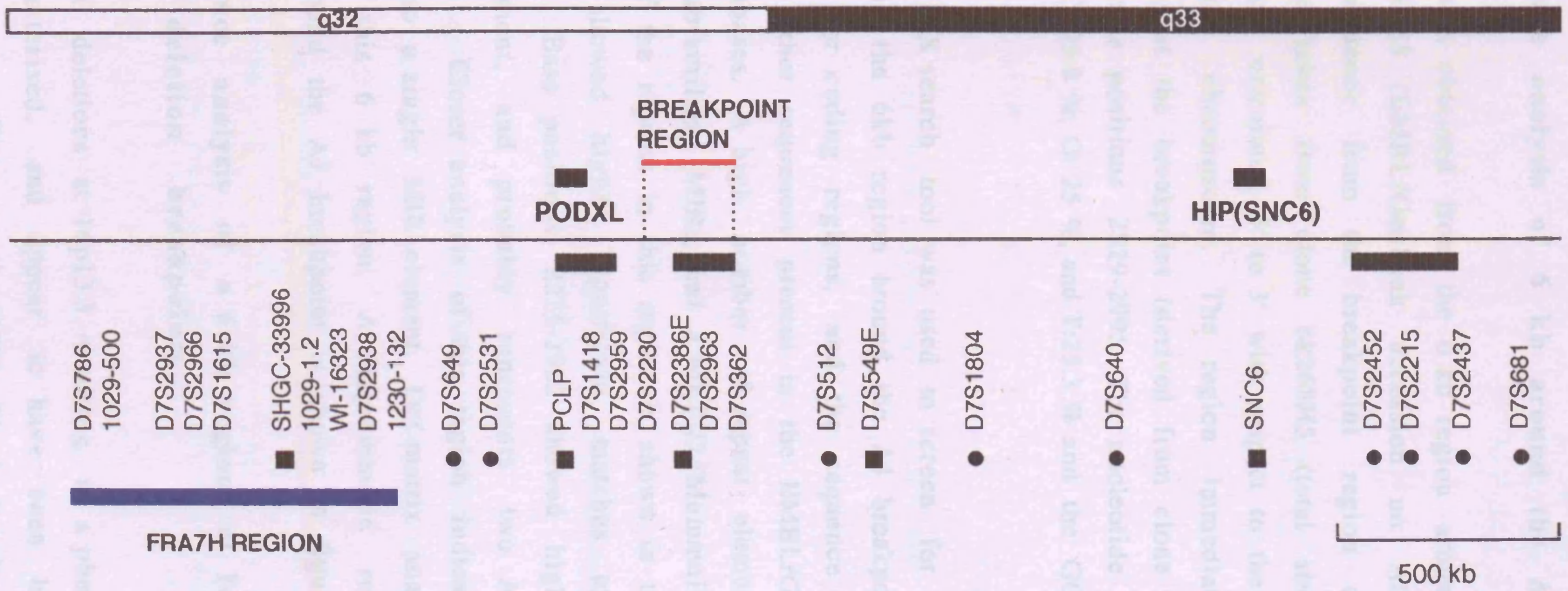


Figure 6.3 Location of the FB336R breakpoint region on the 7q physical map

The FB336R breakpoint region, detected by probe 6a8f, is shown as a red bar (mapping by Dr. S. Scherer, personal communication) and maps between genes PODXL and HIP(SNC6). The breakpoint lies at least 500 kb distal to fragile site region FRA7H which is indicated by a blue bar. Solid black bars indicate groups of markers for which the local order has not been determined. ●denote genetic markers; ■ denote ESTs or genes. Other markers shown are STSs or unique DNA probes. Adapted from figure 1, Mishmar *et al.*, 1998.

been determined, and this region has not yet been sequenced. The subcloned breakpoint region is likely to be located between STS/EST markers, as no hits were identified during database searching, despite the high density of markers mapped in this region.

6.5 Sequence analysis of 6 kb around the AJ breakpoint

Sequence was obtained from the 6 kb region around the AJ breakpoint from BAC clone bk268H5 (EMBL/Genbank accession no. AL008718), representing normal genomic sequence from the breakpoint region of band 22q13.3. Bases 4,102 to 10,102 were taken from clone bk268H5 (total size approximately 231 kb). The sequence was orientated 5′ to 3′ with respect to the telomere repeat array found on the truncated chromosome. The region immediately adjacent to the telomere repeat array at the breakpoint (derived from clone AJ-111b in the AJ library) was located at base positions 2829-2995. The nucleotide composition of the region was A: 25.9 %; C: 23.8 %; G: 25 %, and T:25.3 % and the GC-content of the region was 48.8 %.

Again the NIX search tool was used to screen for putative coding sequences and exons within the 6kb region around the AJ breakpoint. There was no evidence for CpG islands or coding regions, and the sequence did not show any significant matches to other sequences present in the EMBL/Genbank nucleotide, protein, EST or STS databases. A high number of repeat elements were identified, including four *Alu* subfamilies, MIRs, and a MLT1D (Mammalian LTR-Transposon) repeat. A summary of the repeats in this region is shown in table 6.2. Regions 2191-2274 and 2903-2940 showed highly significant matches to a MIR repeat in a reverse orientation. Base positions 2275-2902 showed highly significant matches to an *Alu*-Sx element, and probably represents two Alu elements in a head-to-tail arrangement. Closer analysis of this region indicated that the *Alu* elements had inserted into a single MIR element. Dot matrix analysis did not reveal any other repeats in this 6 kb region. A diagrammatic representation of the sequence features around the AJ breakpoint is shown in figure 6.2.

6.6 Sequence analysis of a 6 kb region at 16p13.3 encompassing three terminal deletion breakpoints

Six terminal deletions at 16p13.3 resulting in a phenotype of α-thalassaemia have been characterised, and appear to have been healed by *de novo* addition of telomere repeats (Flint *et al.*, 1994). These breakpoints have been described in detail in previous chapters. Five of the breakpoints are clustered within a 50 kb

	number of elements	length occupied	percentage of sequence
SINEs:	11	3018 bp	50 %
ALUs	9	2708 bp	45 %
MIRs	2	310 bp	5 %
LINEs:	0		
LINE1	0		
LINE2	0		
LTR elements:	1	329 bp	5 %
MaLRs	1	329 bp	5 %
Retrov.	0		
MER4	0		
DNA elements:	0		
MER1	0		
MER2	0		
Mariners	0		
Total interspersed repeats:		3347 bp	56 %
Small RNA:	0		
Satellites:	0		
Simple repeats:	0		
Low complexity:	0		

Table 6.2 Summary of repeats identified in the 6 kb sequence region around the putative AJ terminal deletion breakpoint

Data shown was obtained using RepeatMasker version 6/16/98db and ProcessRepeats version 06/16/98 via the NIX search tool at the HGMP Human Genome Mapping Centre. Most of the repeats fragmented by insertions or deletions are shown as one element.

region, and breakpoints in patients CMO, IdF and TAT are in close proximity, within a 6 kb region around the embryonic α-like zeta globin gene ζ2 (Proudfoot *et al.*, 1982). The sequence from this 6 kb region was obtained from cosmid clone GG1 (EMBL/Genbank accession number Z84721; Flint and Higgs, unpublished). This cosmid is part of a contig from the tip of 16p which spans 2 Mb of 16p13.3 and contains the α- and ζ-globin genes and ESTs. Bases 13,000 to 19,000 were orientated 5′ to 3′ with respect to the G-rich strand of the novel telomere repeat arrays added at all three breakpoints. The nucleotide composition of the region was A: 24.5 %; C: 25.3 %; G: 30.1 % and T: 20.1 %, and the GC-content of the region was 55.5 %. In these three individuals, it was not possible to determine at which nucleotide the telomere repeats were added, as the 3-4 nucleotides at the breakpoint were identical to, and in register with, the telomere repeat array (Flint *et al.*, 1994). The first possible nucleotide to which telomere repeats were added in individuals CMO, IdF and TAT were located at base positions 434, 3669 and 5296 respectively, shown in figure 6.2. As this region has already been sequenced and characterised, the NIX search confirmed the location of the ζ2-globin gene (EMBL/Genbank accession number M24173) within this region. The ζ2-globin gene has three exons located at base positions 5132-5256 (exon 1), 4068-4247 (exon 2) and 3740-3868 (exon 3). The TAT breakpoint is located 60 bp distal to the start of exon 1 of the ζ2 globin gene, and the IdF breakpoint lies approximately 50 bp proximal to the end of the final exon (exon 3) of the ζ2 globin gene. Intron 1 contains fourteen copies of a simple 14 bp repeat and intron 2 contains 35 copies of a 5 bp repeat with a canonical CGGGG structure (Proudfoot *et al.*, 1982). These intron repeats were not detected by the RepeatMasker program, but were detected using dot matrix analysis. The locations of other repeated regions such as *Alu* elements, L1 repeats and MER repeats were detected by the RepeatMasker program, and are summarised in table 6.3.

6.7 Comparing the 6 kb regions around the AJ, FB336R and 16p13.3 breakpoints

Analysis of the 6 kb regions in the vicinity of the AJ, FB336R and 16p13.3 breakpoints indicated that the only common sequence features were interspersed repeat elements such as *Alu* elements and MIR repeats. The AJ breakpoint lies within very close proximity (approximately 40 bp) to an *Alu*-Sx repeat. The TAT breakpoint lies approximately 700 bp proximal to an *Alu*-S element, 60 bp distal to the start of exon 1 of the ζ2 globin gene and 240 bp distal to the simple 14 bp repeat in intron 1. IdF lies 215 bp proximal to the 5 bp repeat in intron 2 of the ζ2 globin gene, 46 bp proximal to the start of exon 3, and approximately 440 bp distal to an L2 (LINE) element. CMO lies approximately 170 bp distal to an L1 (LINE) repeat and 200

	number of elements	length occupied	percentage of sequence
SINEs:	7	1644 bp	27 %
ALUs	7	1644 bp	27 %
MIRs	0		
LINEs:	2	272 bp	5 %
LINE1	1	119 bp	2 %
LINE2	1	153 bp	3 %
LTR elements:	1	439 bp	7 %
MaLRs	0		
Retrov.	0		
MER4	1	439 bp	7 %
DNA elements:	0		
MER1	0		
MER2	0		
Mariners	0		
Total interspersed repeats:		2355 bp	39 %
Small RNA:	0		
Satellites:	0		
Simple repeats:	0		
Low complexity:	0		

Table 6.3 Summary of repeats identified in the 6 kb sequence region around three 16p13.3 terminal deletion breakpoint

Data shown was obtained using RepeatMasker version 6/16/98db and ProcessRepeats version 06/16/98 via the NIX search tool at the HGMP Human Genome Mapping Centre. Most of the repeats fragmented by insertions or deletions are shown as one element.

bp proximal to an *Alu*-Jo element. The FB336R breakpoint is located 1380 bp distal to an L1 element interspersed with *Alu*s, and 1180 bp proximal to a MIR repeat, slightly further away from repeat regions than the other breakpoints, but still in relatively close proximity. The 6 kb regions around the FB336R, AJ and 16p13.3 breakpoints were compared using the FASTA program, to look for any similar regions or motifs that may have been overlooked using other search tools. The only sequence similarities identified were between related *Alu* elements, as expected.

6.8 Sequence analysis of 60 kb around the AJ breakpoint

The 6 kb of sequence around the putative AJ breakpoint did not contain any notable sequence features other than common repetitive elements. However, further sequence data from this region was available from the EMBL/Genbank database (accession no. AL008718). A 60 kb sequence region was obtained from clone bk268H5 including bases 1-60,000. The bulk of this sequence was from chromosome 22 distal to the breakpoint region, with the putative AJ breakpoint located at base positions 6,931-7,087 at the left end of the BAC clone, orientated 5′ to 3′ with respect to the telomere repeat array on the truncated chromosome. Sequence proximal to this was not contained in this clone. This analysis could potentially lead to the identification of genes in the breakpoint region that may contribute to the phenotype in AJ, if deleted.

The NIX search tool was used to screen for putative coding sequences and exons within the 60kb region encompassing the AJ breakpoint, and the output file is shown in figure 6.4. The region was 52.5 % GC-rich, and two genes were identified. Sequence region 34820-35136 showed significant nucleotide sequence identity (88 %) and protein sequence identity (74 %) to the human 40s ribosomal protein S10 fragment (accession no. U14972/SW:RS10). This region coincided with strong exon predictions from GRAIL/exons and GENEMARK (protein coding exon probability 83 %), and strong single exon gene predictions from Genefinder and Fgenes programs. Predictions for a polyA tail in this region were identified by three independent programs. The start of this single exon gene was located 27,733 bp distal to the breakpoint region.

Sequence region 45163-55096 showed highly significant nucleotide sequence identity (98 % to 100 %), and protein sequence identity (86 % to 98 %), to the uroplakin III mRNA (accession no. AF085808) and protein (accession no. 075631) over six specific regions. These regions coincided with strong exon predictions from five independent programs (Fex, Hexon, MZEF, Genemark and GRAIL/exons),

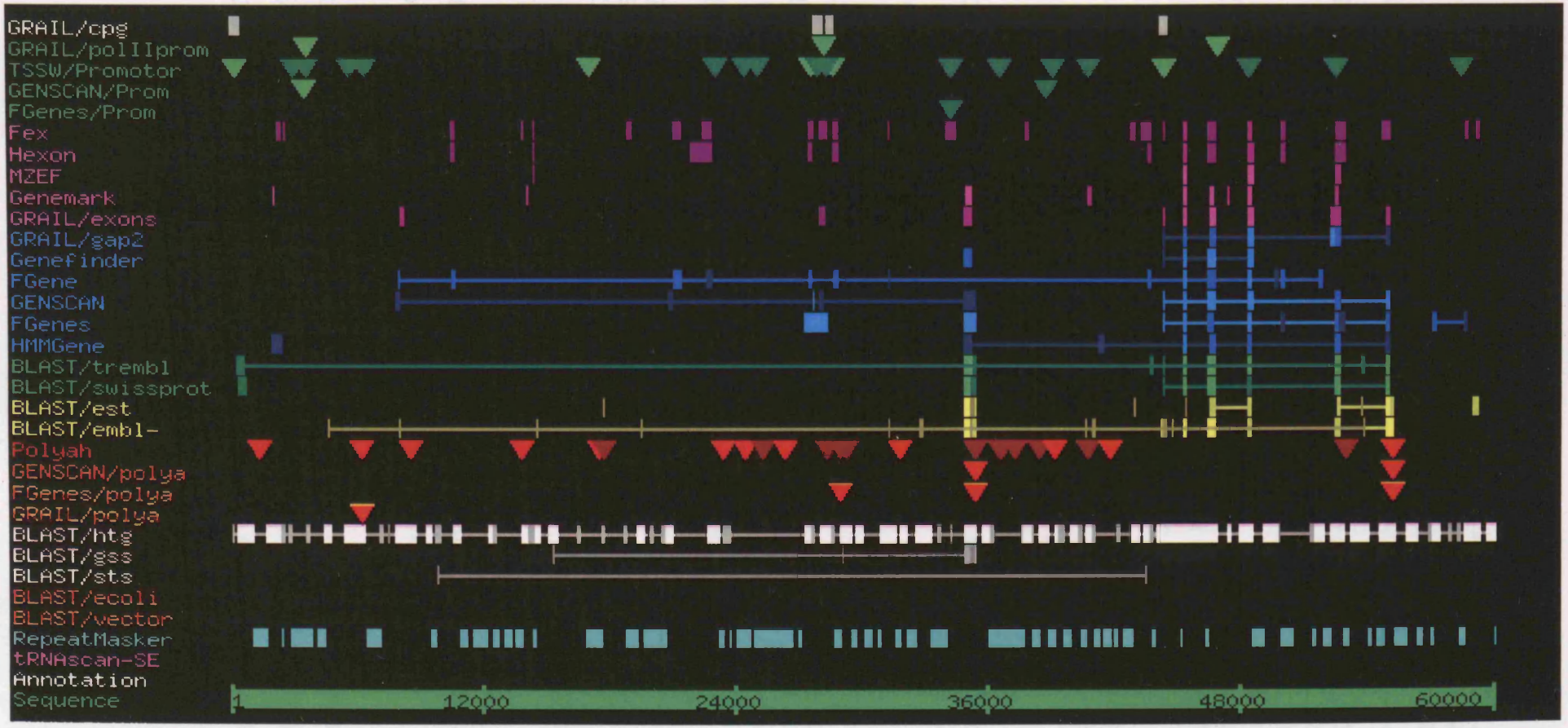


Figure 6.4 NIX search tool output from the 60 kb region around the AJ breakpoint

Sequence analysis was done using the NIX search tool at the HGMP Human Genome Mapping Centre. Programs used are listed down the left hand side of the output file above. The coloured regions represent matches, and brighter regions represent more significant matches. The two genes identified in the region are clearly visible on this output file. The single exon human 40s ribosomal protein S10 fragment is located at approximately 3600 bp (scale indicated by the green bar at the foot of the output file), and is identified by blocks of colour aligned along the same region. The uroplakin III gene has six exons, which can be seen in the region around 44,000 bp to 55,000 bp.

	number of elements	length occupied	percentage of sequence
SINEs:	8	320929 bp	35 %
ALUs	66	18729 bp	31 %
MIRs	17	2200 bp	4 %
LINEs:	21	6955 bp	12 %
LINE1	5	3024 bp	5 %
LINE2	16	3931 bp	7 %
LTR elements:	9	5157 bp	9 %
MaLRs	7	3061 bp	5 %
Retrov.	0		
MER4	1	1614 bp	3 %
DNA elements:	4	923 bp	2 %
MER1	4	923 bp	2 %
MER2	0		
Mariners	0		
Total interspersed repeats:		33964 bp	57 %
Small RNA:	1	33 bp	0.06 %
Satellites:	0		
Simple repeats:	0		
Low complexity:	0		

Table 6.4 Summary of repeats identified in the 60 kb sequence region around the AJ terminal deletion breakpoint

Data shown was obtained using RepeatMasker version 6/16/98db and ProcessRepeats version 06/16/98 via the NIX search tool at the HGMP Human Genome Mapping Centre. Most repeats fragmented by insertions or deletions are shown as one element.

and strong gene predictions from five independent programs (GRAIL/gap2, Fgene, GENSCAN, Genefinder and HMMGene). Strong promoter predictions coincided with the first exon, and polyA predictions coincided with the final exon. The GRAIL/cpg program strongly predicted a CpG island, although it is only 327 bp in length. This is shorter than CpG islands generally associated with most housekeeping and many tissue specific genes, usually 500-1000 bp (Bird *et al.*, 1987). The start of exon 1 was located 37,122 bp distal to the breakpoint region.

Repeats identified in this 60 kb sequence are summarised in table 6.4. A high density of SINEs were identified (including *Alu* and MIR repeats), and interspersed repeats comprised a total of 33,964 bp (56.6 % sequence) within this region.

Discussion

6.9 The isochore structure of the human genome

Human metaphase chromosomes can be defined by three structural regions, Giemsa (G) or Quinacrine (Q) bands, Reverse (R) bands and Centromeric (C) bands. These bands are produced using fluorescent dyes, proteolytic digestion or denaturation (reviewed Comings, 1978; Therman *et al.*, 1986). Most genes map to R-bands (reviewed Gardiner, 1996), which are relatively GC-rich and early replicating, whereas G-bands are relatively AT-rich, late replicating, and thought to be poor in expressed genes (Korenberg and Engels, 1978; Kuhn and Therman, 1986). G and R bands are characterised by distinct families of interspersed repeats. For example, the SINE *Alu* family dominates in R bands, whereas the LINE L1 family dominates in G positive bands (Korenberg and Rykowski, 1988). Alu elements are 56 % GC rich, L1 elements are 58 % AT-rich, and these families can make up 13 % to 18 % of the total DNA within a band. R-bands and R/G-band junctions appear to be predominant sites of exchange processes, for example translocations and sister chromatid exchange (reviewed Morgan and Crossen, 1977). The preferential location of some chromosome exchange processes in R-bands may be related to the clustering of *Alu* elements within these regions, for example α-globin deletions (Nicholls *et al.*, 1987), LDL receptor deletions (Lehrman *et al.*, 1985) and β-globin cluster deletions (Henthorn *et al.*, 1986). It may also be related the open chromatin structure found in these domains, as chromatin is less condensed in transcriptionally active regions.

On average the mammalian genome is 40 % GC-rich, although the distribution is not uniform. The genome is composed of long stretches of DNA (>300 kb) called isochores, homogeneous in base composition. Isochores vary in GC-content from <38 % to >55 %, and consists of five classes: AT-rich L1 and L2, and increasingly GC-rich H1, H2 and H3 (reviewed Gardiner, 1996). Isochores are not uniformly represented in the genome, but do not have a random distribution either, and differ in gene density, CpG dinucleotide density, SINE and LINE distribution and DNA replication timing (reviewed Bernardi, 1989; 1995). Two thirds of the genome is composed of AT-rich L1 and L2 isochore but contains only one third of all genes. The H3 isochores are the most gene-rich regions, representing only 3 % of the genome, but they contain more than one quarter of all genes. L1 and L2 isochores give a pattern similar to G-banding and H1 and H2 isochores give a pattern similar to R-banding. Most proterminal regions contain H3 isochores, although H3 isochores are also found internally (Saccone *et al.*, 1992).

6.10 Sequence features around the AJ, FB336R and 16p13.3 breakpoints

The 6 kb around the AJ (22q13.3) and FB336R (7q32) breakpoints did not yield any putative coding sequences, although 6 kb is a relatively short region. The 6 kb analysed around the three 16p13.3 breakpoints contains the ζ2-globin gene (Proudfoot *et al.*, 1982). All five breakpoint regions studied contained a high density of interspersed repeat elements, in particular SINEs such as *Alu* elements and MIR repeats, and they all map to R-bands, indicating that they lie within relatively GC-rich regions. No other common features were found. The 6 kb regions around FB336R, AJ and 16p13.3 breakpoints are 47 %, 48.8% and 55.5 % GC-rich respectively. These values are high relative to the 40 % genome average and FB336R and AJ are close to the average GC-content calculated for a H2 isochore region (49%; reviewed Gardiner, 1996). However, the AJ breakpoint lies within the terminal band of 22q. Most terminal bands consist of H3 isochore, approximately 53% GC-rich, such as the 16p terminal band 16p13.3 (55.5 %). Analysis of the larger 60 kb region around the AJ breakpoint indicated a GC-content of 53 % as expected for a H3 isochore region, and it is likely that the difference in the GC-content in the shorter sequence reflects differences in GC-content found within isochores.

Sequence characterization of 60 kb around the AJ breakpoint identified two genes, the 40s ribosomal protein S10 fragment, approximately 28 kb distal to the breakpoint, and the uroplakin III gene, approximately 37 kb distal to the breakpoint. Uroplakin III is a membrane protein found only in urothelial cells and is a marker of urothelial differentiation (Yuasa *et al.*, 1998). It seems unlikely that the Uroplakin III gene is associated with the mental retardation and developmental delay seen in AJ, given its function. These genes and their expression have been previously characterised, although they had not previously been mapped to chromosome 22, and it is not known whether these copies are functional or pseudogenes. The region is rich in common interspersed repeats which comprised 57 % of the total sequence. No other sequence features of note were identified.

6.11 Dispersed repeats and chromosome breaks

Most terminal deletion breakpoints isolated to date lie within or in close proximity to an *Alu* element (Flint *et al.*, 1994; discussed in chapter 5). However, all of these breakpoints are found within R-bands, which are by nature GC- and *Alu*-rich regions. It has been suggested that the high incidence of *Alus* (and other repeats) within these terminal breakpoint regions may be merely a consequence of the

Alu-richness of these regions. However, *Alu* elements have been shown to be directly involved in illegitimate recombination events leading to chromosome rearrangement at 16p13.3 (Flint *et al.*, 1996). The 6 kb region around the AJ breakpoint is an *Alu*-rich region (which may be predicted as this breakpoint lies in an R-band). This region contains nine *Alu* elements, and the breakpoint is located only 40 bp distal to an *Alu*. The average *Alu* repeat frequency in the human genome is approximately 1 per 4 kb (Schmid and Jelinek, 1982), although this varies between isochores. In contrast, the 6 kb region around the FB336R breakpoint contains only two *Alu* elements but the breakpoint lies in relatively close proximity (1400 bp distal) to an *Alu*. Other dispersed repeats are present in the FB336R breakpoint region, including MIR elements, an L1 and a MER element.

It is likely that the extensive polymorphism and shared repetitive elements in subterminal regions leads to mispairing and recombination between non-homologous chromosomes, leading to chromosome breakage and in some cases capture of an existing telomere (Flint *et al.*, 1996). This type of chromosome rearrangement may extend to other regions of the genome containing repeated genes or elements that allow homologous or non-homologous recombination. A high density of repeat elements that may be involved in homologous and illegitimate recombination events may predispose towards chromosome breakage, and subsequent loss of the chromosome end in these regions.

6.12 Other sequence regions and chromosome breaks

A number of sequence elements other than interspersed repeats may be involved in recombination and chromosomal rearrangement. Large chromosome deletions, mainly interstitial, tend to be relatively common in regions containing clusters of related genes, such as the β-globin locus (Henthorn *et al.*, 1990), or duplicated genes. Examples of rearrangements within duplicated gene regions include the α-globin gene cluster, which contains repeated α-globin genes α1 and α2, encoding identical polypeptides, an α-globin pseudogene (reviewed Proudfoot *et al.*, 1982), the α-like ζ2 globin gene and the ζ1 pseudogene (Lauer *et al.*, 1980; Proudfoot *et al.*, 1982). Smith-Magenis syndrome (SMS) is one of the most common contiguous microdeletion syndromes, and is caused by interstitial deletions at 17p11.2. The phenotype includes mental retardation, short stature, craniofacial and skeletal abnormalities (Greenberg *et al.*, 1991). Three copies of a low copy number repeat (SMS-REPs) have been identified around the common deletion region, representing a complex repeated gene cluster (Chen *et al.*, 1997). It is thought that one of the frequent mechanisms leading to this microdeletion syndrome is unequal homologous recombination within the SMS-REP gene cluster.

Characterisation of a cryptic rearrangement at 18q21.3 in a patient with 18q- deletion syndrome indicated that the breakpoint was located between two almost identical serine protease inhibitor (serpin) genes SCCA1 and SCCA2 (squamous cell carcinoma antigen) (Katz *et al.*, 1999). These genes are 95 % homologous at the nucleotide level, located less than 10 kb apart and arranged in a head-to-tail orientation (Schneider *et al.*, 1995). No other motifs associated with recombination were identified, and it is possible that recombination occurred in this region due to the repetition of the serpin genes, leading to chromosome breakage and healing by telomere capture (Katz *et al.*, 1999). Initiation of homologous recombination events of this nature in repeated regions may cause chromosome breaks which are then healed by *de novo* telomere addition, leading to the generation of terminal deletions.

No evidence for the presence of repeated genes was found in the FB336R breakpoint region. However, only 6 kb around the FB336R breakpoint was studied, and it is likely that genes are found outside this region, relatively close to the breakpoint. Sequence (60 kb) around the AJ breakpoint was examined, and two genes, Uroplakin III and the 40s ribosomal protein S10 fragment, mapped to this region distal to the breakpoint. Most of the sequence examined mapped distal to the breakpoint and it is possible these genes are repeated, or have pseudogene copies in the regions proximal to the breakpoint.

Fragile sites are inherited chromosomal anomalies, and appear as constrictions in the chromosome, susceptible to breakage and rearrangement under certain experimental conditions *in vitro* (Hecht and Sutherland, 1985; Sutherland and Richards, 1995). These sites may also be linked with chromosomal rearrangements and deletions (Jones *et al.*, 1994). Common fragile sites are thought to be part of normal chromosome structure, but the only one giving a definite clinical phenotype is the rare fragile site FRAXA at Xq27.3, causing fragile X syndrome, a common form of inherited mental retardation (Brown and Jenkins, 1992; Giraud *et al.*, 1976). The molecular basis of fragile X syndrome is the expansion of the CCG-trinucleotide repeat in the 5′ untranslated region of the FMR1 (fragile-X mental retardation 1) gene associated with FRAXA (Brown and Jenkins, 1992; Kremer *et al.*, 1991), associated methylation of an adjacent CpG island (Oberle *et al.*, 1991; Vincent *et al.*, 1991) and reduction in the transcription of FMR1 (Devys *et al.*, 1993; Pieretti *et al.*, 1991). Autosomal fragile sites have not yet been associated with genetic disease in carriers, although there is evidence that nine autosomal fragile sites are found at higher levels in mentally retarded individuals than in the general population (Sutherland, 1982).

An autosomal fragile site has been linked with one case of Jacobsen syndrome, usually associated with loss of part of the long arm of chromosome 11 (generally 11q23 to 11qter), and a clinical phenotype of severe mental retardation and dysmorphic features (Jacobsen, 1973; Schinzel *et al.*, 1977). The breakpoint in one Jacobsen patient localises to band 11q23.3. A rare inherited fragile site, FRA11B (Voullaire *et al.*, 1987) maps to the same 100 kb region. It was not known whether the affected child had inherited the FRA11B chromosome, but it is possible that the break occurred within the maternally inherited FRA11B chromosome during early development. The truncated chromosome appears to have been stabilised by addition of a telomere or telomere-like sequence (Jones *et al.*, 1994), although sequence across the breakpoint was not characterised.

Common fragile sites FRA7G and FRA7H have been recently identified in bands 7q31.2 and 7q32.3 respectively (Mishmar *et al.*, 1998), and the FB336R breakpoint maps to the same band (7q32) as the FRA7H site. A region of approximately 160 kb spanning the FRA7H common fragile site has been sequenced (EMBL/Genbank accession no. AF017104; Mishmar *et al.*, 1998) and shows unusual chromatin organisation, suggesting under-condensation of this region. The FRA7H sequence is A+T-rich (58 %), and is composed of 13.1 % SINEs, 13.8 % LINEs, 5 % LTRs, and appears to be gene poor. No other extended repeats such as CGG were found, in contrast to rare fragile sites. Two pseudogenes were found within the FRA7H region (zinc finger protein ZNF131 and histone H4) representing several integration events which may be associated with the instability of this region (Mishmar *et al.*, 1998). These sequence characteristics, such as open chromatin, may extend through the entire region around the fragile site and predispose towards chromosome breakage.

The inheritance of fragile sites may be a common mechanism for chromosome breakage and deletion syndromes, although evidence for chromosome breakage around other known fragile sites has not yet been found (Jones *et al.*, 1994). The FB336R breakpoint maps at least 500 kb distal to the FRA7H fragile site, and it is therefore unlikely that the fragile site played a direct role in this chromosome break. The relationship between fragile sites and terminal deletions is as yet unclear, and more detailed investigations are required.

In summary, the only shared sequence features around FB336R (7q32), AJ (22q13.3) and the three 16p13.3 breakpoints were common interspersed repeat elements. All five breakpoints were located in R-bands, and all were in terminal R-bands except FB336R. The 16p13.3 breakpoints are located in the α-globin locus, which contains duplicated genes that can be associated with chromosome

breakage. The AJ and 16p breakpoints were found within relatively *Alu*-rich regions. The FB336R breakpoint region appeared to contain a lower density of *Alu* repeats in comparison, but did contain copies of other dispersed repeats. Repeat elements, particularly *Alu*s, are though to play a role in illegitimate recombination events, which can cause chromosome rearrangement and deletion. However, it is not clear whether these repeats have a role in promoting healing by *de novo* telomere addition. No other features that have been associated with chromosome breakage were identified in close proximity to the FB336R and AJ breakpoints.

It has been suggested that healing by telomerase requires a short region of complementarity to the telomerase RNA template immediately at the breakpoint in order for telomerase to bind and add repeats (Flint *et al.*, 1994; discussed in chapter 5). The presence of G-rich sequence in the immediate vicinity of the breakpoint may allow recognition of the chromosome break by telomerase, but does not appear to be a strict requirement. TTAGGG repeats were not found around any of the breakpoints characterised to date. It is not known whether telomerase has long-range sequence requirements, or prefers some sequences over others when adding repeats. However, no sequence features of note were identified within the 6 kb region around the AJ, FB336R and 16p13.3 breakpoints, or within 60 kb around the AJ breakpoint. Short (6 kb) and long-range (60 kb) sequence analysis around a number of other terminal deletion breakpoints healed by *de novo* telomere addition, including breakpoints located in terminal and interstitial bands, R- and G-bands, is required before any firm conclusion can be drawn.

Chapter 7

FINAL DISCUSSION AND FUTURE WORK

Discussion

Work presented in this thesis has been described in detail in the relevant chapters and is summarised in this final chapter.

The main objective of this project was to use a telomere-anchored PCR based strategy to isolate breakpoints from patients with suspected terminal deletions, which have been healed by *de novo* telomere addition. It is highly likely that this type of healing event is mediated by the enzyme telomerase, although it is not known at what frequency such events occur *in vivo*, and whether telomerase has specific sequence requirements for recognising a broken chromosome end and adding repeats. The existing strategy was unsuccessful in isolating terminal deletion breakpoints from six patients, but this work led to isolation of two novel chromosome ends, and a number of improvements to the ordered array library screening step. A further aim was to modify the existing strategy to include an additional enrichment step for telomere repeat arrays when generating libraries containing telomere-junction clones. The main findings were:

1) A low frequency polymorphic telomere (CB0001-14d) in 2.4 % of the Caucasian population. This telomere is unusual in that the adjacent sequence shows strong similarity to a subterminal repeat found at 22q, 4q, 10q and 4p which is not normally located immediately adjacent to a telomere. In a few individuals, a copy of this subterminal repeat is truncated by an array of telomere repeats. This probably creates a telomere-junction at a location more internal than in most individuals.

2) Isolation of a novel subterminal repeat (CB0001-24c) found adjacent to a telomere on chromosomes 7, 11 and 17. The chromosomal distribution of this subterminal repeat may be polymorphic.

3) Compilation of a sequence database, comprising subterminal sequences from telomere-junction clones, representing normal chromosome ends. This database can be queried using a FASTA search tool, avoiding unnecessary investigation of previously isolated telomere-junction clones.

4) Generation of additional probes to detect subterminal repeats found at normal telomere-junctions on a number of chromosomes.

5) Development of variant repeat probes to detect variant repeats TGAGGG, TTGGGG and TCAGGG within a repeat array, thus eliminating the majority of normal

chromosome ends. In conjunction with subterminal repeat probes, the variant repeat probes greatly reduced unnecessary sequencing of normal chromosome ends containing these variant repeats.

Attempts to modify the telomere-anchored PCR and cloning steps to yield a uniform high cloning efficiency of telomere-junctions were unsuccessful. This work demonstrated that the strategy used to generate ordered array libraries was optimal within the technical limits. An additional step to further enrich for telomere repeat arrays was developed, based on filter hybridisation selection. When used in conjunction with size fractionation of *Mbo*I-digested DNA and telomere-anchored PCR, filter hybridisation selection was shown to increase enrichment for telomere repeat arrays, specifically those containing only TTAGGG repeats. Filter hybridisation also resulted in the exclusion of a number of normal chromosome ends containing variant repeat types prior to cloning. This modified strategy was used to generate an ordered array library from an individual with a terminal deletion at 22q13.3, from which a candidate terminal deletion breakpoint was isolated. This clone is thought to represent the breakpoint due to the following features:

1) The telomere-adjacent sequence did not show homology to previously identified subterminal repeats, and was mapped to chromosome 22.

2) The non-telomeric sequence at the putative breakpoint was found adjacent to an array of telomere repeats in individual AJ, but not 87 unrelated individuals, or the parents of the patient, indicating that it is a novel telomere repeat array.

3) The telomere-adjacent region showed 96.2 % sequence identity to a BAC clone mapped to 22q13.3, the region to which the breakpoint was initially mapped by RFLP analysis (Nesslinger *et al.*, 1994).

4) The telomere repeat array appears to contain TTAGGG repeats only, without variants, strongly suggesting that it is a newly synthesised telomere repeat array.

The mechanism of healing at this candidate breakpoint is unknown, but experimental evidence indicates healing by direct addition of telomere repeats to non-telomeric sequence, most likely by telomerase. The mechanism by which telomerase recognises a broken chromosome end and adds repeats is not clear. A 6 kb sequence region around the putative breakpoint at 22q13.3 was compared with sequence from around a terminal deletion breakpoint at 7q32, and three clustered terminal deletion breakpoints at 16p13.3 (Flint *et al.*, 1994), which also appear to have been healed by direct addition of a telomere. The only sequence feature common to these breakpoint regions was a high density of dispersed repeats, particularly *Alu* elements.

Chromosome breakage and healing

There are two principal mechanisms by which broken chromosomes can be stabilised: the direct addition of a newly synthesised telomere onto non-telomeric DNA, probably by telomerase; or capture of an existing telomere, either by recombination or DNA polymerisation using the existing telomere as a template. There is evidence that both these mechanisms contribute to the stabilisation of truncated chromosomes. A number of examples of chromosome healing by telomerase have been described, including six breakpoints at 16p13.3 (Flint *et al.*, 1994), one at 7q32 (N. J. Royle, unpublished data), one at 22q13.3 (Wong *et al.*, 1997) and the putative AJ breakpoint described in this thesis, also at 22q13.3. The frequency at which this type of healing event occurs is unknown. Examples of telomere capture resulting in terminal deletion have been found at 16p13.3, via illegitimate recombination between two related *Alu* elements (Flint *et al.*, 1996), and at 18q, via illegitimate recombination with a satellite III region (Katz *et al.*, 1999). It is likely that telomere capture results in the transfer of subterminal or repetitive sequence along with the telomere repeat array, which is likely to contain variant or degenerate repeats, distinguishing this type of healing event from *de novo* telomere addition. The telomere-anchored PCR screening strategy is designed to isolate novel telomeres only, and cannot be used to isolate truncated chromosomes where subterminal sequence remains at a chromosome end, or where the telomere contains variant repeats.

All terminal deletion breakpoints isolated to date lie in GC-rich R-bands, and all but FB336R are located in terminal bands. It is possible that chromosome breakage occurs more frequently in these bands. R-bands are gene rich regions (reviewed Gardiner, 1996) where active transcription can occur. Consequently these regions have an 'open chromatin' structure some, or all, of the time. R-bands are also characterised by distinct families of dispersed repeats, in particular *Alu* elements. *Alu* repeats have been shown to be involved in homologous and illegitimate recombination events leading to interstitial deletions (Nicholls *et al.*, 1987; Henthorn *et al.*, 1986; Lehrman *et al.*, 1985) and terminal deletions (Flint *et al.*, 1996). The abundance of these repeats and the 'open chromatin' conformation in R-bands may predispose towards chromosome breakage in these regions via recombination. Chromosome breaks may then be stabilised by the addition of a new telomere.

Chromosome healing events at 16p13.3 are perhaps more easily identified than healing events on other chromosomes because of the associated phenotype of α-thalassaemia. The lack of a recognisable clinical phenotype associated with

terminal deletions of other chromosome ends has prevented detailed analysis of such rearrangements. It is thought that this type of deletion is the basis for a number of cases of idiopathic mental retardation (Flint *et al.*, 1995). Common polymorphic subterminal repeats found at many chromosome ends make identification of individual chromosome ends difficult, particularly if subterminal sequence remains at a truncated chromosome end. Mental retardation is a common feature of terminal microdeletions, probably as loss of only one haplo-insufficient gene in a developmental pathway is likely to have significant consequences for the developing fetus. Telomerase activity has been detected in the germline and in the blastocyst stage of development (Wright *et al.*, 1996b), generating a short window during which chromosome healing by telomerase may occur.

It is not clear whether telomerase has specific sequence requirements for recognising a broken chromosome end and adding telomere repeats. Telomerase has been shown to recognise the sequence at one of the breakpoints at 16p13.3 as a substrate and add repeats *in vitro* (Morin, 1991). This strongly suggests that telomerase was able to recognise the broken chromosome end *in vivo* and add a new telomere. It was thought that telomerase requires a G-rich region, or telomere-like repeats near the breakpoint in order to recognise a DNA strand, bind and begin to add repeats (reviewed Blackburn, 1992a). *In vitro*, telomerase shows preference for some oligonucleotides as substrates for telomere repeat addition. Human telomerase can elongate oligonucleotides with 3′ ends without extensive complementarity to the template region (Harrington and Greider, 1991; Morin, 1991) and it may only require a few nucleotides complementary to the template region of the telomerase RNA component to bind and add repeats. However, the presence of G-rich sequence greatly enhances the efficiency of initiation and elongation (Harrington and Greider, 1991). *Tetrahymena* telomerase can also utilise primers consisting of non-telomeric pBR322 vector sequences at the 3′ end and TTGGGG repeats at the 5′ end. The pBR322 sequence is cleaved off and repeats added to the TTGGGG sequence with the same efficiency as the addition of repeats to $(TTGGGG)_4$ primers (Harrington and Greider, 1991). Similar cleavage by human telomerase, followed by nucleotide addition by telomerase, or another enzyme, may lead to the generation of 'orphan' nucleotides at terminal deletion breakpoints.

The 5′ nucleotide composition of oligonucleotide substrates strongly affects the efficiency of *de novo* telomere addition to non-telomeric primers *in vitro* by *Tetrahymena* telomerase (Wang and Blackburn, 1997). This suggests that telomerase may not use the RNA template for initial recognition and binding, but may carry out substrate recognition at the second binding site (anchor site), which binds residues 5′ to the site of telomere repeat addition. Alternatively, 5′

sequences may be recognised by DNA-protein interactions with a protein component of telomerase (Collins and Greider, 1993). A protein-DNA recognition mechanism may be relevant to direct telomere addition *in vivo*, and may not require telomere-like or G-rich motifs at the 3′ end of the substrate sequence. In *Tetrahymena*, the sequences at the sites of telomere addition during macronuclear development do not act as substrates for telomerase *in vitro*, suggesting that other factors interact *in vivo* to assist in programmed chromosome healing in these organisms (Spangler *et al.*, 1988), which may or may not be present in other species. Comparison of sequences at the nine terminal deletion breakpoints previously described suggests that there are no strict sequence requirements for telomerase binding and repeat addition, apart from a possible minimal 2-4 bp region of complementarity to the RNA template, and a free 3′ DNA end. G residues at the breakpoint may help binding and stabilisation of telomerase (Flint *et al.*, 1994). The presence of 2-3 nucleotides complementary to the RNA template at the breakpoints is also sufficient for healing by telomerase when chromosomes are broken by integration of artificial contracts in human cell lines (Barnett *et al.*, 1993; Murnane and Yu, 1993). No TTAGGG-like repeats were found in the vicinity of the breakpoints isolated to date, and G-rich motifs were found immediately proximal to the breakpoint in only a few cases, although presence of such motifs may increase healing efficiency. However, two of the breakpoints do not show any complementarity to the telomerase RNA template, and appear to have 'orphan' nucleotides inserted at the breakpoint. The putative AJ breakpoint described in this thesis has ten 'orphan' nucleotides inserted between the normal chromosome 22 sequence and the telomere repeat array. These nucleotides may have been generated during the healing event by the action of telomerase and other enzymes, or may indicate a more complex healing event. Alternatively, they may represent sequence polymorphism in this region.

Evidence described in previous chapters indicates that telomerase does not require specific sequence motifs to recognise broken chromosome ends or add a new telomere. The isolation of a number of other terminal deletion breakpoints that appear to have been healed by this mechanism, from both terminal and interstitial bands, may allow conclusions to be drawn about the requirements of telomerase, if any, for healing broken chromosomes, and the frequency at which this type of healing event occurs *in vivo*.

Future Work

The putative AJ terminal deletion breakpoint

The telomere repeat array at the putative AJ breakpoint generates telomere-anchored PCR products of up to 2.3 kb, strongly suggesting it is located at a chromosome end. Sensitivity of this repeat array to *Bal*31 digestion should confirm a terminal location. The apparent insertion of the ten 'orphan' nucleotides between the normal chromosome 22 sequence and the start of the *de novo* telomere repeat array may be explained by a common sequence polymorphism, or sequencing errors in the unfinished BAC clone representing the normal genomic copy of this region of chromosome 22. Sequencing across this region in AJ's parents, or a number of other individuals, may resolve these differences, or establish whether they have arisen during the healing event. These orphan nucleotides may indicate a more complex healing event, although it is probably not possible to ascertain whether such an event took place due to the short length of the 'orphan' sequence. The discrepancy in the exact location of the breakpoint, initially determined by RFLP and densitometric analysis of polymorphic markers (Nesslinger *et al.*, 1994), should be resolved. Densitometric analysis may have given misleading results and could be repeated in order to resolve this discrepancy. Alternatively, PCR amplification of polymorphic markers in the breakpoint region in AJ and parents (KJ and FJ) could be used to detect loss of heterozygosity, providing that the parents are both heterozygous for a given marker, and that they do not have alleles of the same size.

Using filter hybridisation selection to isolate other terminal deletion breakpoints

Generation of telomere-junction libraries, using the modified strategy including the filter hybridisation selection enrichment step, from a number of other individuals with suspected terminal deletions should result in the isolation of additional breakpoints healed by the addition of a novel telomere. This modified strategy could also be used to attempt to isolate the NT terminal deletion breakpoint, which has an *Mbo*I-digestion fragment of 3.5 kb. This would determine whether the filter hybridisation step efficiently enriches for longer amplicons with long stretches of telomere-adjacent sequence, or whether it preferentially enriches for telomere repeat arrays with only short telomere-adjacent sequence.

Sequence analysis at other terminal deletion breakpoints

Nine terminal deletion breakpoints that appear to have been healed by the direct addition of telomere repeats to a broken end by telomerase have been isolated to date. Six of these breakpoints localise to 16p13.3, two localise to 22q13.3, and one to band 7q32. At present it does not appear that telomerase has any specific sequence requirements for generating *de novo* telomeres *in vivo*. Comparison of short and long range sequence data from around other terminal deletion breakpoints is required before any firm conclusion can be drawn.

In Conclusion

In this thesis I have described the isolation of a candidate terminal deletion breakpoint at 22q13.3, from a patient with a phenotype including mental retardation. The structure of the putative breakpoint is similar to previously isolated terminal deletion breakpoints that appear to have been healed by direct addition of telomere repeats to non-telomeric DNA, consistent with healing by telomerase.

Probe	Number of positives	% of total number of clones	% of telomere positive clones
Tel +	258	34	
Tsk37	62	8	24
Tsk46	12	2	5
Tsk8 (Xp/Yp)	6	1	2
E-F (7q/12q)	51	7	20
PGSEGB	100	13	39
CB0001-24c	12	2	5
TelG	164	22	64
TelJ	142	19	55
TelK	24	3	9

Appendix I. Results of screening the ROBAIIA ordered array library with subterminal and variant repeat probes

The library contained a total number of 752 clones. Probes used for screening are shown in the far left-hand column. Subsequent columns indicate the number of positive clones in the library for each probe, the percentage of the total number of clones positive for each probe and the percentage of telomere positive clones that were also positive with each probe.

References

Allshire, R.C., Dempster, M. and Hastie, N.D. (1989) Human telomeres contain at least three types of G-rich repeat distributed non-randomly. *Nucleic Acids Res.*, **17**, 4611-4627.

Allshire, R.C., Gosden, J.R., Cross, S.H., Cranston, G., Rout, D., Sugawara, N., Szostak, J.W., Fantes, P.A. and Hastie, N.D. (1988) Telomeric repeat from *T. thermophila* cross hybridizes with human telomeres. *Nature*, **332**, 656-659.

Allsopp, R.C. (1996) Models of initiation of replicative senescence by loss of telomeric DNA. *Exp. Gerontol.*, **31**, 235-243.

Allsopp, R.C. and Harley, C.B. (1995) Evidence for a critical telomere length in senescent human fibroblasts. *Expt. Cell Res.*, **219**, 130-136.

Allsopp, R.C., Vaziri, H., Patterson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W. and Harley, C.B. (1992) Telomere length predicts replicative capacity of human fibroblasts. *Proc. Natl. Acad. Sci. USA*, **89**, 10114-10118.

Altherr, M.R., Bengtsson, U., Elder, F.F., Ledbetter, D.H., Wasmuth, J.J., McDonald, M.E., Gusella, J.F. and Greenberg, F. (1991) Molecular confirmation of Wolf-Hirschhorn syndrome with a subtle translocation of chromosome 4. *Am. J. Hum. Genet.*, **49**, 1235-1242.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403-410.

Aparicio, O.M., Billington, B.L. and Gottschling, D.E. (1991) Modifiers of position effect are shared between telomeric and silent mating-type loci in *S. cerevisiae*. *Cell*, **66**, 1279-1287.

Armour, J.A. (1990) The isolation and characterisation of human minisatellites. PhD thesis. University of Leicester.

Armour, J.A. and Jeffreys, A.J. (1991) STS for minisatellite MS607 (D22S163). *Nucleic Acids Res.*, **19**, 3158.

Armour, J.A., Neumann, R., Gobert, S. and Jeffreys, A.J. (1994) Isolation of human simple repeat loci by hybridization selection. *Hum. Mol. Genet.*, **3**, 599-565.

Armour, J.A., Povey, S., Jeremiah, S. and Jeffreys, A.J. (1990) Systematic cloning of human minisatellites from ordered array charomid libraries. *Genomics*, **8**, 501-512.

Armour, J.A., Wong, Z., Wilson, V., Royle, N.J. and Jeffreys, A.J. (1989) Sequences flanking the repeat arrays of human minisatellites: association with tandem and dispersed repeat elements. *Nucleic Acids Res.*, **17**, 4925-4935.

Bacchetti, S. and Counter, C.M. (1995) Telomeres and telomerase in human cancer (Review). *Int. J. Oncol.*, **7**, 423-432.

Baird, D.M. (1996) The analysis of variation at the human Xp:Yp telomere. PhD thesis. University of Leicester.

Baird, D.M., Coleman, J., Rosser, Z.H. and Royle, N.J. (1999) High levels of sequence polymorphism and linkage disequilibrium at the telomere of 12q; implications for telomere biology and human evolution. *In prep.*

Baird, D.M., Jeffreys, A.J. and Royle, N.J. (1995) Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human Xp/Yp telomere. *EMBO J.*, **14**, 5433-5443.

Baird, D.M. and Royle, N.J. (1997) Sequences from higher primates orthologous to the human Xp/Yp telomere junction region reveal gross rearrangements and high levels of divergence. *Hum. Mol. Genet.*, **6**, 2291-2299.

Barinaga, M. (1997) Cell biology - Cells count proteins to keep their telomeres in line. *Science*, **275**, 928-928.

Barnes, W.M. (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl. Acad. Sci. USA*, **91**, 2216-2220.

Barnett, M.A., Buckle, V.J., Evans, E.P., Porter, A.C., Rout, D., Smith, A.G. and Brown, W.R. (1993) Telomere directed fragmentation of mammalian chromosomes. *Nucleic Acids Res.*, **21**, 27-36.

Bates, G.P., MacDonald, M.E., Baxendale, S., Sedlacek, Z., Youngman, S., Romano, D., Whaley, W.L., Allitto, B.A., Poustka, A., Gusella, J.F. and *et al.* (1990) A yeast artificial chromosome telomere clone spanning a possible location of the Huntington disease gene. *Am. J. Hum. Genet.*, **46**, 762-775.

Batzer, M.A., Schmid, C.W. and Deininger, P.L. (1993) Evolutionary analyses of repetitive DNA sequences. *Methods Enzymol.*, **224**, 213-232.

Baxendale, S., MacDonald, M.E., Mott, R., Francis, F., Lin, C., Kirby, S.F., James, M., Zehetner, G., Hummerich, H., Valdes, J. and *et al.* (1993) A cosmid contig and high resolution restriction map of the 2 megabase region containing the Huntington's disease gene. *Nat. Genet.*, **4**, 181-186.

Bayne, R.A., Broccoli, D., Taggart, M.H., Thomson, E.J., Farr, C.J. and Cooke, H.J. (1994) Sandwiching of a gene within 12 kb of a functional telomere and alpha satellite does not result in silencing. *Hum. Mol. Genet.*, **3**, 539-546.

Bender, M.A., Preston, R.J., Leonard, R.C., Pyatt, B.E. and Gooch, P.C. (1989) Chromosomal aberration and sister-chromatid exchange frequencies in peripheral blood lymphocytes of a large human population sample. II. Extension of age range. *Mutat. Res.*, **212**, 149-154.

Bengtsson, U., Altherr, M.R., Wasmuth, J.J. and Winokur, S.T. (1994) High resolution fluorescence *in situ* hybridization to linearly extended DNA visually maps a tandem repeat associated with facioscapulohumeral muscular dystrophy immediately adjacent to the telomere of 4q. *Hum. Mol. Genet.*, **3**, 1801-1805.

Bergold, P.J., Campbell, G.R., Littau, V.C. and Johnson, E.M. (1983) Sequence and hairpin structure of an inverted repeat series at termini of the *Physarum* extrachromosomal rDNA molecule. *Cell*, **32**, 1287-1299.

Bernardi, G. (1989) The isochore organization of the human genome. *Annu. Rev. Genet.*, **23**, 637-661.

Bernardi, G. (1995) The human genome: organization and evolutionary history. *Annu. Rev. Genet.*, **29**, 445-476.

Bianchi, A., Smith, S., Chong, L., Elias, P. and de Lange, T. (1997) TRF1 is a dimer and bends telomeric DNA. *EMBO J.*, **16**, 1785-1794.

Biessmann, H., Champion, L.E., O'Hair, M., Ikenaga, K., Kasravi, B. and Mason, J.M. (1992) Frequent transpositions of *Drosophila melanogaster* HeT-A transposable elements to receding chromosome ends. *EMBO J.*, **11**, 4459-4469.

Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Buffer gradient gels and 35S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA*, **80**, 3963-3965.

Bilaud, T., Brun, C., Ancelin, K., Koering, C.E., Laroche, T. and Gilson, E. (1997) Telomeric localization of TRF2, a novel human telobox protein. *Nat. Genet.*, **17**, 236-239.

Bilaud, T., Koering, C.E., Binet-Brasselet, E., Ancelin, K., Pollice, A., Gasser, S.M. and Gilson, E. (1996) The telobox, a Myb-related telomeric DNA binding motif found in proteins from yeast, plants and human. *Nucleic Acids Res.*, **24**, 1294-1303.

Bird, A.P., Taggart, M.H., Nicholls, R.D. and Higgs, D.R. (1987) Non-methylated CpG-rich islands at the human alpha-globin locus: implications for evolution of the alpha-globin pseudogene. *EMBO J.*, **6**, 999-1004.

Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.*, **7**, 1513-1523.

Blackburn, E.H. (1992a) Telomerases. *Annu. Rev. Biochem.*, **61**, 113-129.

Blackburn, E.H. (1992b) Telomeres and their synthesis. *The Harvey Lectures*, Series **86**, 1-18.

Blackburn, E.H. (1984) The molecular structure of centromeres and telomeres. *Annu. Rev. Biochem.*, **53**, 163-194.

Blackburn, E.H., Budarf, M.L., Challoner, P.B., Cherry, J.M., Howard, E.A., Katzen, A.L., Pan, W.C. and Ryan, T. (1983) DNA termini in ciliate macronuclei. *Cold Spring Harb. Symp. Quant. Biol.*, **47**, 1195-1207.

Blackburn, E.H. and Chiou, S.S. (1981) Non-nucleosomal packaging of a tandemly repeated DNA sequence at termini of extrachromosomal DNA coding for ribosomal RNA in *Tetrahymena. Proc. Nat. Acad. Sci. USA*, **78**, 2263-2267.

Blackburn, E.H. and Gall, J.G. (1978) A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in *Tetrahymena. J. Mol. Biol.*, **120**, 33-53.

Blasco, M.A., Lee, H.W., Hande, M.P., Samper, E., Lansdorp, P.M., DePinho, R.A. and Greider, C.W. (1997) Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell*, **91**, 25-34.

Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C.P., Morin, G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S. and Wright, W.E. (1998) Extension of life-span by introduction of telomerase into normal human cells. *Science*, **279**, 349-352.

Boulton, S.J. and Jackson, S.P. (1996) Identification of a *Saccharomyces cerevisiae* Ku80 homologue: Roles in DNA double strand break rejoining and in telomeric maintenance. *Nucleic Acids Res.*, **24**, 4639-4648.

Bourgain, F.M. and Katinka, M.D. (1991) Telomeres inhibit end to end fusion and enhance maintenance of linear DNA molecules injected into the *Paramecium primaurelia* macronucleus. *Nucleic Acids Res.*, **19**, 1541-1547.

Brigati, C., Kurtz, S., Balderes, D., Vidali, G. and Shore, D. (1993) An essential yeast gene encoding a TTAGGG repeat-binding protein. *Mol. Cell. Biol.*, **13**, 1306-1314.

Broccoli, D., Chong, L., Oelmann, S., Fernald, A.A., Marziliano, N., van Steensel, B., Kipling, D., Le Beau, M.M. and de Lange, T. (1997a) Comparison of the human and mouse genes encoding the telomeric protein, TRF1: chromosomal localization, expression and conserved protein domains. *Hum. Mol. Genet.*, **6**, 69-76.

Broccoli, D. and Cooke, H. (1993) Aging, healing, and the metabolism of telomeres. *Am. J. Hum. Genet.*, **52**, 657-660.

Broccoli, D., Smogorzewska, A., Chong, L. and deLange, T. (1997b) Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nat. Genet.*,*17*, 231-235.

Broccoli, D., Young, J.W. and de Lange, T. (1995) Telomerase activity in normal and malignant hematopoietic cells. *Proc. Natl. Acad. Sci. USA*, **92**, 9082-9086.

Brown, W.R.A. (1989) Molecular cloning of human telomeres in yeast. *Nature*, **338**, 774-776.

Brown, W.R.A., Mackinnon, P.J., Villasante, A., Spurr, N., Buckle, V.J. and Dobson, M.J. (1990) Structure and polymorphism of human telomere-associated DNA. *Cell*, **63**, 119-132.

Brown, W.T. and Jenkins, E.C. (1992) The fragile X syndrome. *Mol. Genet. Med.*, **2**, 39-66.

Bryan, T.M., Englezou, A., Gupta, J., Bacchetti, S. and Reddel, R.R. (1995) Telomere elongation in immortal human cells without detectable telomerase activity. *EMBO J.*, **14**, 4240-4248.

Buard, J. and Jeffreys, A.J. (1997) Big, bad minisatellites. *Nat. Genet.*, **15**, 327-328.

Buluwela, L., Forster, A., Boehm, T. and Rabbitts, T.H. (1989) A rapid procedure for colony screening using nylon filters. *Nucleic Acids Res.*, **17**, 452.

Campisi, J. (1996) Replicative senescence - an old lives tale. *Cell*, **84**, 497-500.

Cardenas, M.E., Bianchi, A. and de Lange, T. (1993) A *Xenopus* egg factor with DNA-binding properties characteristic of terminus-specific telomeric proteins. *Genes Dev.*, **7**, 883-894.

Chan, C.S. and Tye, B.K. (1983) Organization of DNA sequences and replication origins at yeast telomeres. *Cell*, **33**, 563-573.

Chen, K.S., Manian, P., Koeuth, T., Potocki, L., Zhao, Q., Chinault, A.C., Lee, C.C. and Lupski, J.R. (1997) Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nat. Genet.*, **17**, 154-163.

Cheng, J., Smith, C.L. and Cantor, C.R. (1990) Structural and transcriptional analysis of a human subtelomeric repeat. *Nucleic Acids Res.*, **19**, 149-154.

Chong, L., Vansteensel, B., Broccoli, D., Erdjumentbromage, H., Hanish, J., Tempst, P. and Delange, T. (1995) A human telomeric protein. *Science*, **270**, 1663-1667.

Chong, S.S., Tanigami, A., Roschke, A.V. and Ledbetter, D.H. (1996) 14-3-3 epsilon has no homology to LIS1 and lies telomeric to it on chromosome 17p13.3 outside the Miller-Dieker syndrome chromosome region. *Genome Res.*, **6**, 735-741.

Church, G.M. and Gilbert, W. (1984) Genomic Sequencing. *Proc. Natl. Acad. Sci. USA*, **81**, 1991-1995.

Cohen, J. (1994) 'Long PCR' leaps into larger DNA sequences. *Science*, **263**, 1564-1565.

Cohn, M. and Blackburn, E.H. (1995) Telomerase in yeast. *Science*, **269**, 396-400.

Coleman, J. (1998) The analysis of variation at human autosomal telomeres. PhD thesis. University of Leicester.

Coleman, J., Baird, D.M. and Royle, N.J. (1999) The plasticity of human telomeres demonstrated by a hypervariable telomere-repeat array that is located on some copies of 16p and 16q. *Hum. Mol. Genet.*, submitted.

Coleman, M.J., McHale, M.T., Arnau, J., Watson, A. and Oliver, R.P. (1993) Cloning and characterisation of telomeric DNA from *Cladosporium fulvum*. *Gene*, **132**, 67-73.

Collins, K. and Greider, C.W. (1993) *Tetrahymena* telomerase catalyzes nucleolytic cleavage and nonprocessive elongation. *Genes Dev.*, **7**, 1364-1376.

Collins, K., Kobayashi, R. and Greider, C.W. (1995) Purification of *Tetrahymena* telomerase and cloning of genes encoding the two protein components of the enzyme. *Cell*, **81**, 677-686.

Comings, D.E. (1978) Methods and mechanisms of chromosome banding. *Methods Cell. Biol.*, **17**, 115-132.

Connor, J.M. and Ferguson-Smith, M.A. Essential Medical Genetics. Third Edition. Blackwell Science Ltd.

Conrad, M.N., Wright, J.H., Wolf, A.J. and Zakian, V.A. (1990) RAP1 protein interacts with yeast telomeres in vivo: overproduction alters telomere structure and decreases chromosome stability. *Cell*, **63**, 739-750.

Cooper, J.P., Nimmo, E.R., Allshire, R.C. and Cech, T.R. (1997) Regulation of telomere length and function by a Myb-domain protein in fission yeast. *Nature*, **385**, 744-747.

Coulson, A., Sulston, J., Brenner, S. and Karn, J. (1986) Toward a physical map of the nematode *Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA*, **83**, 7821-7825.

Counter, C.M., Avilion, A.A., Lefeuvre, C.E., Stewart, N.G., Greider, C.W., Harley, C.B. and Bacchetti, S. (1992) Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J.*, **11**, 1921-1929.

Counter, C.M., Botelho, F.M., Wang, P., Harley, C.B. and Bacchetti, S. (1994a) Stabilization of short telomeres and telomerase activity accompany immortalization of Epstein-Barr virus-transformed human B lymphocytes. *J. Virol.*, **68**, 3410-3414.

Counter, C.M., Gupta, J., Harley, C.B., Leber, B. and Bacchetti, S. (1995) Telomerase activity in normal leukocytes and in hematologic malignancies. *Blood*, **85**, 2315-2320.

Counter, C.M., Hahn, W.C., Wei, W.Y., Caddle, S.D., Beijersbergen, R.L., Lansdorp, P.M., Sedivy, J.M. and Weinberg, R.A. (1998) Dissociation among *in vitro* telomerase activity, telomere maintenance, and cellular immortalization. *Proc. Natl. Acad. Sci. USA*, **95**, 14723-14728.

Counter, C.M., Hirte, H.W., Bacchetti, S. and Harley, C.B. (1994b) Telomerase activity in human ovarian carcinoma. *Proc. Natl. Acad. Sci. USA*, **91**, 2900-2904.

Counter, C.M., Meyerson, M., Eaton, E.N. and Weinberg, R.A. (1997) The catalytic subunit of yeast telomerase. *Proc. Natl. Acad. Sci. USA*, **94**, 9202-9207.

Craig, J.M. and Bickmore, W.A. (1994) The distribution of CpG islands in mammalian chromosomes. *Nat. Genet.*, **7**, 376-382.

Cross, S.H., Allshire, R.C., McKay, S.J., McGill, N.I. and Cooke, H.J. (1989) Cloning of human telomeres by complementation in yeast. *Nature*, **338**, 771-774.

Cross, S., Lindsey, J., Fantes, J., McKay, S., McGill, N. and Cooke, H. (1990) The structure of a subterminal repeated sequence present on many human chromosomes. *Nucleic Acids Res.*, **18**, 6649-6657.

Coulson, A., Sulston, J., Brenner, S. and Karn, J. (1986) Toward a physical map of the nematode *Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA.*, **83**, 7821-7825.

de la Sena, C., Chowdhary, B.P. and Gustavsson, I. (1995) Localization of the telomeric (TTAGGG)n sequences in chromosomes of some domestic animals by fluorescence *in situ* hybridization. *Hereditas*, **123**, 269-274.

de Lange, T. (1992) Human telomeres are attached to the nuclear matrix. *EMBO J.*, **11**, 717-724.

de Lange, T., Shiue, L., Myers, R.M., Cox, D.R., Naylor, S.L., Killery, A.M. and Varmus, H.E. (1990) Structure and variability of human chromosome ends. *Mol. Cell. Biol.*, **10**, 518-527.

Denhardt, D.T. (1966) A membrane-filter technique for the detection of complementary DNA. *Biochem. Biophys. Res. Commun.*, **23**, 641-646.

Devereux, J., Haeberli, P. and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.*, **12**, 387-395.

Devys, D., Lutz, Y., Rouyer, N., Bellocq, J.P. and Mandel, J.L. (1993) The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. *Nat. Genet.*, **4**, 335-340.

Dice, J.F. (1987) Molecular determinants of protein half-lives in eukaryotic cells. *FASEB J.*, **1**, 349-357.

Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T.P., Bowden, D.W., Smith, D.R., Lander, E.S. and et al. (1987) A genetic linkage map of the human genome. *Cell*, **51**, 319-337.

Drwinga, H.L., Toji, L.H., Kim, C.H., Greene, A.E. and Mulivor, R.A. (1993) NIGMS human/rodent somatic cell hybrid mapping panels 1 and 2. *Genomics*, **16**, 311-314.

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

Emery, H.S. and Weiner, A.M. (1981) An irregular satellite sequence is found at the termini of the linear extrachromosomal rDNA in *Dictyostelium discoideum*. *Cell*, **26**, 411-419.

Evans, A.E., D'Angio, G.J. and Randolph, J. (1971) A proposed staging for children with neuroblastoma. Children's cancer study group A. *Cancer*, **27**, 374-378.

Farr, C.J. (1996) Mammalian telomeres and chromosome fragmentation. *Cell Dev. Biol.*, **7**, 41-48.

Farr, C., Fantes, J., Goodfellow, P. and Cooke, H. (1991) Functional reintroduction of human telomeres into mammalian cells. *Proc. Natl. Acad. Sci. USA*, **88**, 7006-7010.

Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.*, **132**, 6-13.

Feinberg, A.P. and Vogelstein, B. (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Addendum. *Anal. Biochem.*, **137**, 266-267.

Feng, J.L., Funk, W.D., Wang, S.S., Weinrich, S.L., Avilion, A.A., Chiu, C.P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J.H., Le, S.Y., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W. and Villeponteau, B. (1995) The RNA component of human telomerase. *Science*, **269**, 1236-1241.

Flint, J., Bates, G.P., Clark, K., Dorman, A., Willingham, D., Roe, B.A., Micklem, G., Higgs, D.R. and Louis, E.J. (1997) Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Hum. Mol. Genet.* **6**, 1305-1313

Flint, J., Craddock, C.F., Villegas, A., Bentley, D.P., Williams, H.J., Galanello, R., Cao, A., Wood, W.G., Ayyub, H. and Higgs, D.R. (1994) Healing of broken human chromosomes by the addition of telomeric repeats. *Am. J. Hum. Genet.*, **55**, 505-512.

Flint, J., Rochette, J., Craddock, C.F., Dode, C., Vignes, B., Horsley, S.W., Kearney, L., Buckle, V.J., Ayyub, H. and Higgs, D.R. (1996) Chromosomal stabilization by a subtelomeric rearrangement involving two closely-related *Alu* elements. *Hum. Mol. Genet.*, **5**, 1163-1169.

Flint, J., Wilkie, A.O.M., Buckle, V.J., Winter, R.M., Holland, A.J. and McDermid, H.E. (1995) The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. *Nat. Genet.*, **9**, 132-140.

Forney, J.D. and Blackburn, E.H. (1988) Developmentally controlled telomere addition in wild-type and mutant paramecia. *Mol. Cell. Biol.*, **8**, 251-258.

Frezal, J. and Schinzel, A. (1991) Report of the committee on clinical disorders, chromosome aberrations and uniparental disomy. *Cytogenet. Cell Genet.*, **58**, 986-1052.

Gardiner, K. (1996) Base composition and gene distribution: critical patterns in mammalian genome organization. *Trends Genet.*, **12**, 519-524.

Garvik, B., Carson, M. and Hartwell, L. (1995) Single-stranded DNA arising at telomeres in cdc13 mutants may constitute a specific signal for the RAD9 checkpoint. *Mol. Cell. Biol.*, **15**, 6128-6138.

Gieselmann, V., Fluharty, A.L., Tonnesen, T. and Von Figura, K. (1991) Mutations in the arylsulfatase A pseudodeficiency allele causing metachromatic leukodystrophy. *Am. J. Hum. Genet.*, **49**, 407-413.

Gilley, D. and Blackburn, E.H. (1996) Specific RNA residue interactions required for enzymatic functions of *Tetrahymena* telomerase. *Mol Cell Biol*, 16, 66-75.

Gilson, E., Mullers, I., Sogo, J., Laroche, T. and Gasser, S.M. (1994) Rap1 stimulates single-strand to double-strand association of yeast telomeric DNA - implications for telomere-telomere interactions. *Nucleic Acids Res.*, 22, 5310-5320.

Gilson, E., Roberge, M., Giraldo, R., Rhodes, D. and Gasser, S.M. (1993) Distortion of the DNA double helix by Rap1 at silencers and multiple telomeric binding sites. *J. Mol. Biol.*, **231**, 293-310.

Giraud, F., Ayme, S., Mattei, J.F. and Mattei, M.G. (1976) Constitutional chromosomal breakage. *Hum. Genet.*, **34**, 125-136.

Giraudeau, F., Aubert, D., Young, I., Horsley, S., Knight, S., Kearney, L., Vergnaud, G. and Flint, J. (1997) Molecular-cytogenetic detection of a deletion of 1p36.3. *J. Med. Genet.*, **34**, 314-317.

Gottschling, D.E. (1992) Telomere-proximal DNA in *Saccharomyces cerevisiae* is refractory to methyltransferase activity *in vivo*. *Proc. Natl. Acad. Sci. USA*, **89**, 4062-4065.

Gottschling, D.E., Aparicio, O.M., Billington, B.L. and Zakian, V.A. (1990) Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell*, **63**, 751-762.

Gottschling, D.E. and Cech, T.R. (1984) Chromatin structure of the molecular ends of *Oxytricha* macronuclear DNA - phased nucleosomes and a telomeric complex. *Cell*, **38**, 501-510.

Gottschling, D.E. and Zakian, V.A. (1986) Telomere Proteins - specific recognition and protection of the natural termini of *Oxytricha* macronuclear DNA. *Cell*, **47**, 195-205.

Greaves, M. (1996) Is telomerase activity in cancer due to selection of stem cells and differentiation arrest. *Trends Genet.*, **12**, 127-128.

Greenberg, F., Guzzetta, V., Montes de Oca-Luna, R., Magenis, R.E., Smith, A.C., Richter, S.F., Kondo, I., Dobyns, W.B., Patel, P.I. and Lupski, J.R. (1991) Molecular analysis of the Smith-Magenis syndrome: a possible contiguous gene syndrome associated with del(17)(p11.2). *Am. J. Hum. Genet.*, **49**, 1207-1218.

Greenwell, P.W., Kronmal, S.L., Porter, S.E., Gassenhuber, J., Obermaier, B. and Petes, T.D. (1995) Tel1, a gene involved in controlling telomere length in *Saccharomyces cerevisiae*, is homologous to the human ataxia telangiectasia gene. *Cell*, **82**, 823-829.

Greider, C.W. (1990) Telomeres, telomerase and senescence. *Bioessays*, **12**, 363-369.

Greider, C.W. and Blackburn, E.H. (1985) Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell*, **43**, 405-413.

Greider, C.W. and Blackburn, E.H. (1987) The telomere terminal transferase of *Tetrahymena* is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell*, **51**, 887-898.

Greider, C.W. and Blackburn, E.H. (1989) A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. *Nature*, **337**, 331-337.

Guerrini, A.M., Camponeschi, B., Ascenzioni, F., Piccolella, E. and Donini, P. (1993) Subtelomeric as well as telomeric sequences are lost from chromosomes in proliferating B lymphocytes. *Hum. Mol. Genet.*, **2**, 455-460.

Gurrieri, F., Trask, B.J., Vandenengh, G., Krauss, C.M., Schinzel, A., Pettenati, M.J., Schindler, D., Dietzband, J., Vergnaud, G., Scherer, S.W., Tsui, L.C. and Muenke, M. (1993) Physical mapping of the holoprosencephaly critical region on chromosome 7q36. *Nat. Genet.*, **3**, 247-251.

Halvorsen, T.L., Leibowitz, G. and Levine, F. (1999) Telomerase activity is sufficient to allow transformed cells to escape from crisis. *Mol. Cell. Biol.*, **19**, 1864-1870.

Harle-Bachor, C. and Boukamp, P. (1996) Telomerase activity in the regenerative basal layer of the epidermis inhuman skin and in immortal and carcinoma-derived skin keratinocytes. *Proc. Natl. Acad. Sci. USA*, **93**, 6476-6481.

Harley, C.B. (1991) Telomere Loss - Mitotic clock or genetic time momb. *Mutat. Res.*, **256**, 271-282.

Harley, C.B., Futcher, A.B. and Greider, C.W. (1990) Telomeres shorten during ageing of human fibroblasts. *Nature*, **345**, 458-460.

Harley, C.B., Kim, N.W., Prowse, K.R., Weinrich, S.L., Hirsch, K.S., West, M.D., Bacchetti, S., Hirte, H.W., Counter, C.M., Greider, C.W., Piatyszek, M.A., Wright, W.E. and Shay, J.W. (1994) Telomerase, cell immortality, and cancer. *Cold Spring Harb. Symp. Quant. Biol.*, **59**, 307-315.

Harrington, L., Hull, C., Crittenden, J. and Greider, C. (1995) Gel shift and UV cross-linking analysis of *Tetrahymena* telomerase. *J. Biol. Chem.*, **270**, 8893-8901.

Harrington, L., McPhail, T., Mar, V., Zhou, W., Oulton, R., Bass, M.B., Arruda, I. and Robinson, M.O. (1997) A mammalian telomerase-associated protein. *Science, 275*, 973-977.

Harrington, L.A. and Greider, C.W. (1991) Telomerase primer specificity and chromosome healing. *Nature*, **353**, 451-454.

Harris, P.C. and Higgs, D.R. (1993) Structure of the terminal region of the short arm of chromosome 16. *Regional Physical Mapping*, **5**, 107-134. Cold Spring Harbour Laboratory Press.

Harris, P.C. and Thomas, S. (1992) Length polymorphism of the subtelomeric regions of both the short (p) and long arms (q) of chromosome 16. *Cytogenet. Cell Genet.*, **60**, 171-172.

Hastie, N.D., Dempster, M., Dunlop, M.G., Thompson, A.M., Green, D.K. and Allshire, R.C. (1990) Telomere reduction in human colorectal carcinoma and with aging. *Nature*, **346**, 866-868.

Hayflick, L. (1965) The limited *in vitro* lifetime of human diploid cell strains. *Expt. Cell Res.*, **37**, 614-636.

Hayflick, L. and Moorehead, P.S. (1961) The serial cultivation of human diploid cells. *Expt. Cell Res.*, **25**, 585-621.

Hecht, F. and Sutherland, G.R. (1985) Detection of fragile sites on human chromosomes. *Clin. Genet.*, **28**, 95-96.

Henderson, E., Hardin, C.C., Walk, S.K., Tinoco, I., Jr. and Blackburn, E.H. (1987) Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell*, **51**, 899-908.

Henderson, E.R. and Blackburn, E.H. (1989) An overhanging 3' terminus is a conserved feature of telomeres. *Mol. Cell. Biol.*, **9**, 345-348.

Henikoff, S. (1992) Position effect and related phenomena. *Curr. Opin. Genet. Dev.*, **2**, 907-912.

Henthorn, P.S., Mager, D.L., Huisman, T.H. and Smithies, O. (1986) A gene deletion ending within a complex array of repeated sequences 3' to the human beta-globin gene cluster. *Proc. Natl. Acad. Sci. USA*, **83**, 5194-5198.

Henthorn, P.S., Smithies, O. and Mager, D.L. (1990) Molecular analysis of deletions in the human beta-globin gene cluster: deletion junctions and locations of breakpoints. *Genomics*, **6**, 226-237.

Herrera, E., Samper, E. and Blasco, M.A. (1999) Telomere shortening in mTR-/- embryos is associated with failure to close the neural tube. *EMBO J.*, **18**, 1172-1181.

Higgs, D.R., Vickers, M.A., Wilkie, A.O.M., Pretorius, I.M., Jarman, A.P. and Weatherall, D.J. (1989) A review of the molecular genetics of the human alpha-globin gene cluster. *Blood*, **73**, 1081-1104.

Hiyama, E., Hiyama, K., Yokoyama, T., Ichikawa, T. and Matsuura, Y. (1992) Length of telomeric repeats in neuroblastoma: correlation with prognosis and other biological characteristics. *Jpn. J. Cancer Res.*, **83**, 159-164.

Hiyama, E., Hiyama, K., Yokoyama, T., Matsuura, Y., Piatyszek, M.A. and Shay, J.W. (1995) Correlating telomerase activity levels with human neuroblastoma outcomes. *Nat. Med.*, **1**, 249-255.

Holt, S.E., Shay, J.W. and Wright, W.E. (1996a) Refining the telomere-telomerase hypothesis of aging and cancer. *Nat. Biotechnol.*, **14**, 836-839.

Holt, S.E., Wright, W.E. and Shay, J.W. (1996b) Regulation of telomerase activity in immortal cell lines. *Mol. Cell. Biol.*, **16**, 2932-2939.

Horowitz, H., Thorburn, P. and Haber, J.E. (1984) Rearrangements of highly polymorphic regions near telomeres of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **4**, 2509-2517.

Horvath, M.P., Schweiker, V.L., Bevilacqua, J.M., Ruggles, J.A. and Schultz, S.C. (1998) Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA. *Cell*, **95**, 963-974.

Hsiao, R., Sharma, H.W., Ramakrishnan, S., Keith, E. and Narayanan, R. (1997) Telomerase activity in normal human endothelial cells. *Anticancer Res.*, **17**, 827-832.

IJdo, J.W., Baldini, A., Ward, D.C., Reeders, S.T. and Wells, R.A. (1991) Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proc Natl Acad Sci USA*, **88**, 9051-9055.

IJdo, J.W., Lindsay, E.A., Wells, R.A. and Baldini, A. (1992) Multiple variants in subtelomeric regions of normal karyotypes. *Genomics*, **14**, 1019-1025.

Jacobsen, P. (1973) 18p-Syndrome. Deletion of the short arm of chromosome 18. *Ugeskr. Laeger*, **135**, 17.

Jeanpierre, M. (1994) Human satellites 2 and 3. *Ann. Genet.*, **37**, 163-171.

Jeffreys, A.J. (1987) Highly variable minisatellites and DNA fingerprints. *Biochem. Soc. Trans.*, **15**, 309-317.

Jeffreys, A.J., Allen, M.J., Armour, J.A., Collick, A., Dubrova, Y., Fretwell, N., Guram, T., Jobling, M., May, C.A., Neil, D.L. and et al. (1995) Mutation processes at human minisatellites. *Electrophoresis*, **16**, 1577-1585.

Jeffreys, A.J. and Neumann, R. (1997) Somatic mutation processes at a human minisatellite. *Hum. Mol. Genet.*, **6**, 129-132; 134-126.

Jeffreys, A.J., Tamaki, K., MacLeod, A., Monckton, D.G., Neil, D.L. and Armour, J.A. (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nat. Genet.*, **6**, 136-145.

Jeffreys, A.J., Wilson, V., Neumann, R. and Keyte, J. (1988) Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. *Nucleic Acids Res.*, **16**, 10953-10971.

Jiang, X.R., Jimenez, G., Chang, E., Frolkis, M., Kusler, B., Sage, M., Beeche, M., Bodnar, A.G., Wahl, G.M., Tlsty, T.D. and Chiu, C.P. (1999) Telomerase expression in human somatic cells does not induce changes associated with a transformed phenotype. *Nat. Genet.*, **21**, 111-114.

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. *Hum. Mol. Genet.*, **3**, 2123-2130.

Jurka, J. (1989) Subfamily structure and evolution of the human L1 family of repetitive sequences. *J. Mol. Evol.*, **29**, 496-503.

Jurka, J. (1990) Novel families of interspersed repetitive elements from the human genome. *Nucleic Acids Res.*, **18**, 137-141.

Jurka, J. and Milosavljevic, A. (1991) Reconstruction and analysis of human *Alu* genes. *J. Mol. Evol.*, **32**, 105-121.

Jurka, J. and Smith, T. (1988) A fundamental division in the *Alu* family of repeated sequences. *Proc. Natl. Acad. Sci. USA*, **85**, 4775-4778.

Kalaitsidaki, M., Cox, T., Chakravarti, A. and Antonarakis, S.E. (1992) Cloning and linkage mapping of three polymorphic tetranucleotide (TAAA)n repeats on human chromosome 21. *Genomics*, **14**, 1071-1075.

Katinka, M.D. and Bourgain, F.M. (1992) Interstitial telomeres are hotspots for illegitimate recombination with DNA molecules injected into the macronucleus of *Paramecium primaurelia*. *EMBO J.*, **11**, 725-732.

Katz, S.G., Schneider, S.S., Bartuski, A., Trask, B.J., Massa, H., Overhauser, J., Lalande, M., Lansdorp, P.M. and Silverman, G.A. (1999) An 18q- syndrome breakpoint resides between the duplicated serpins SCCA1 and SCCA2 and arises via a cryptic rearrangement with satellite III DNA. *Hum. Mol. Genet.*, **8**, 87-92.

Kennedy, B.K., Austriaco, N.R., Jr. and Guarente, L. (1994) Daughter cells of *Saccharomyces cerevisiae* from old mothers display a reduced life span. *J. Cell Biol.*, **127**, 1985-1993.

Kermouni, A., Van Roost, E., Arden, K.C., Vermeesch, J.R., Weiss, S., Godelaine, D., Flint, J., Lurquin, C., Szikora, J.P., Higgs, D.R. and et al. (1995) The IL-9 receptor gene (IL9R): genomic structure, chromosomal localization in the pseudoautosomal region of the long arm of the sex chromosomes, and identification of IL9R pseudogenes at 9qter, 10pter, 16pter, and 18pter. *Genomics*, **29**, 371-382.

Kilian, A., Bowtell, D.D., Abud, H.E., Hime, G.R., Venter, D.J., Keese, P.K., Duncan, E.L., Reddel, R.R. and Jefferson, R.A. (1997) Isolation of a candidate human telomerase catalytic subunit gene, which reveals complex splicing patterns in different cell types. *Hum. Mol. Genet.*, **6**, 2011-2019.

Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. and Shay, J.W. (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science*, **266**, 2011-2015.

Kipling, D. (1995) Telomerase: immortality enzyme or oncogene? *Nat. Genet.*, **9**, 104-106.

Kipling, D. and Cooke, H.J. (1990) Hypervariable ultra-long telomeres in mice. *Nature*, **347**, 400-402.

Kirk, K.E. and Blackburn, E.H. (1995) An unusual sequence arrangement in the telomeres of the germ-line micronucleus in *Tetrahymena thermophila*. *Genes Dev.*, **9**, 59-71.

Kirk, K.E., Harmon, B.P., Reichardt, I.K., Sedat, J.W. and Blackburn, E.H. (1997) Block in anaphase chromosome separation caused by a telomerase template mutation. *Science*, **275**, 1478-1481. .

Klemm, U., Muller-Esterl, W. and Engel, W. (1991) Acrosin, the peculiar sperm-specific serine protease. *Hum. Genet.*, **87**, 635-641.

Klobutcher, L.A., Gygax, S.E., Podoloff, J.D., Vermeesch, J.R., Price, C.M., Tebeau, C.M. and Jahn, C.L. (1998) Conserved DNA sequences adjacent to chromosome fragmentation and telomere addition sites in *Euplotes crassus*. *Nucleic Acids Res.* ***26***, 4230-4240.

Klobutcher, L.A., Swanton, M.T., Donini, P. and Prescott, D.M. (1981) All gene-sized DNA molecules in four species of Hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proc. Natl. Acad. Sci. USA*, **78**, 3015-3019.

Knight, S.J.L., Horsley, S.W., Regan, R., Lawrie, N.M., Maher, E.J., Cardy, D.L.N., Flint, J. and Kearney, L. (1997) Development and clinical application of an innovative fluorescence *in situ* hybridization technique which detects submicroscopic rearrangements involving telomeres. *Eur. J. Hum. Genet.*, **5**, 1-8.

Korenberg, J.R. and Engels, W.R. (1978) Base ratio, DNA content, and quinacrine-brightness of human chromosomes. *Proc. Natl. Acad. Sci. USA*, **75**, 3382-3386.

Korenberg, J.R. and Rykowski, M.C. (1988) Human genome organization: *Alu*, lines, and the molecular structure of metaphase chromosome bands. *Cell*, **53**, 391-400.

Kremer, E.J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S.T., Schlessinger, D., Sutherland, G.R. and Richards, R.I. (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science*, **252**, 1711-1714.

Kuhn, E.M. and Therman, E. (1986) Cytogenetics of Bloom's syndrome. *Cancer Genet. Cytogenet.*, **22**, 1-18.

Kurtz, S. and Shore, D. (1991) RAP1 protein activates and silences transcription of mating-type genes in yeast. *Genes Dev.*, **5**, 616-628.

Kuwano, A., Ledbetter, S.A., Dobyns, W.B., Emanuel, B.S. and Ledbetter, D.H. (1991) Detection of deletions and cryptic translocations in Miller-Dieker syndrome by *in situ* hybridization. *Am. J. Hum. Genet.*, **49**, 707-714.

Kyrion, G., Boakye, K.A. and Lustig, A.J. (1992) C-terminal truncation of RAP1 results in the deregulation of telomere size, stability, and function in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **12**, 5159-5173.

Kyrion, G., Liu, K., Liu, C. and Lustig, A.J. (1993) RAP1 and telomere structure regulate telomere position effects in *Saccharomyces cerevisiae*. *Genes Dev.*, **7**, 1146-1159.

Lamb, J., Harris, P.C., Lindenbaum, R.H. and Reeders, S.T. (1989) Detection of breakpoints in submicroscopic chromosomal translocations, illustrating an important mechanism for genetic disease. *Lancet*, **2**, 819-824.

Lamb, J., Harris, P.C., Wilkie, A.O.M., Wood, W.G., Dauwerse, J.H.G. and Higgs, D.R. (1993) *De novo* truncation of chromosome 16p and healing with (TTAGGG)n in the alpha-thalassemia mental-retardation syndrome (ATR-16). *Am. J. Hum. Genet.*, **52**, 668-676.

Lansdorp, P.M., Verwoerd, N.P., Vanderijke, F.M., Dragowska, V., Little, M.T., Dirks, R.W., Raap, A.L. and Tanke, H.J. (1996) Heterogeneity in telomere length of human chromosomes. *Hum. Mol. Genet.*, **5**, 685-691.

Lauer, J., Shen, C.K. and Maniatis, T. (1980) The chromosomal arrangement of human alpha-like globin genes: sequence homology and alpha-globin gene deletions. *Cell*, **20**, 119-130.

Ledbetter, D.H. (1992) Cryptic translocations and telomere integrity. *Am. J. Hum. Genet.*, **51**, 451-456.

Lee, H.W., Blasco, M.A., Gottlieb, G.J., Horner, J.W., 2nd, Greider, C.W. and DePinho, R.A. (1998) Essential role of mouse telomerase in highly proliferative organs. *Nature*, **392**, 569-574.

Lee, M.S. and Blackburn, E.H. (1993) Sequence-specific DNA primer effects on telomerase polymerization activity. *Mol. Cell. Biol.*, **13**, 6586-6599.

Lehrman, M.A., Goldstein, J.L., Russell, D.W. and Brown, M.S. (1987) Duplication of seven exons in LDL receptor gene caused by *Alu-Alu* recombination in a subject with familial hypercholesterolemia. *Cell*, **48**, 827-835.

Lehrman, M.A., Schneider, W.J., Sudhof, T.C., Brown, M.S., Goldstein, J.L. and Russell, D.W. (1985) Mutation in LDL receptor: *Alu-Alu* recombination deletes exons encoding transmembrane and cytoplasmic domains. *Science*, **227**, 140-146.

Lejnine, S., Makarov, V.L. and Langmore, J.P. (1995) Conserved nucleoprotein structure at the ends of vertebrate and invertebrate chromosomes. *Proc Natl Acad Sci USA*, **92**, 2393-2397.

Lendvay, T.S., Morris, D.K., Sah, J., Balasubramanian, B. and Lundblad, V. (1996) Senescence mutants of *Saccharomyces cerevisiae* with a defect in telomere replication identify three additional EST genes. *Genetics*, **144**, 1399-1412.

Levis, R.W., Ganesan, R., Houtchens, K., Tolar, L.A. and Sheen, F.M. (1993) Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell*, **75**, 1083-1093.

Li, B. and Lustig, A.J. (1996) A novel mechanism for telomere size control in *Saccharomyces cerevisiae*. *Genes Dev*, **10**, 1310-1326.

Li, L.W., Lejnine, S., Makarov, V. and Langmore, J.P. (1998) *In vitro* and *in vivo* reconstitution and stability of vertebrate chromosome ends. *Nucleic Acids Res.*, **26**, 2908-2916.

Lin, J.J. and Zakian, V.A. (1995) An *in vitro* assay for *Saccharomyces* telomerase requires EST1. *Cell*, **81**, 1127-1135.

Lin, J.J. and Zakian, V.A. (1996) The *Saccharomyces* CDC13 protein is a single-strand TG_{1-3} telomeric DNA- binding protein *in vitro* that affects telomere behavior *in vivo*. *Proc. Natl. Acad. Sci. USA*, **93**, 13760-13765.

Lingner, J. and Cech, T.R. (1996) Purification of telomerase from *Euplotes aediculatus*: requirement of a primer 3' overhang. *Proc. Natl. Acad. Sci. USA*, **93**, 10712-10717.

Lingner, J., Hendrick, L.L. and Cech, T.R. (1994) Telomerase RNAs of different ciliates have a common secondary structure and a permuted template. *Genes Dev.*, **8**, 1984-1998.

Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. and Cech, T.R. (1997) Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science*, **276**, 561-567.

Liu, Z. and Tye, B. (1991) A yeast protein that binds to vertebrate telomeres and conserved yeast telomeric junctions. *Genes Dev.*, **5**, 49-59.

Loidl, J. (1990) The initiation of meiotic chromosome pairing: the cytological view. *Genome*, **33**, 759-778.

Louis, E.J. and Haber, J.E. (1990) Mitotic recombination among subtelomeric Y' repeats in *Saccharomyces cerevisiae*. *Genetics*, **124**, 547-559.

Louis, E.J. and Haber, J.E. (1992) The structure and evolution of subtelomeric Y' repeats in *Saccharomyces cerevisiae*. *Genetics*, **131**, 559-574.

Louis, E.J., Naumova, E.S., Lee, A., Naumov, G. and Haber, J.E. (1994) The chromosome end in yeast: its mosaic nature and influence on recombinational dynamics. *Genetics*, **136**, 789-802.

Lu, X. and Lane, D.P. (1993) Differential induction of transcriptionally active p53 following UV or ionizing radiation - defects in chromosome instability syndromes. *Cell*, **75**, 765-778.

Luderus, M.E., van Steensel, B., Chong, L., Sibon, O.C., Cremers, F.F. and de Lange, T. (1996) Structure, subnuclear distribution, and nuclear matrix association of the mammalian telomeric complex. *J. Cell Biol.*, **135**, 867-881.

Lue, N.F. and Xia, J. (1998) Species-specific and sequence-specific recognition of the dG-rich strand of telomeres by yeast telomerase. *Nucleic Acids Res.*, **26**, 1495-1502.

Luke, S. and Verma, R.S. (1993) Telomeric repeat (TTAGGG)n sequences of human chromosomes are conserved in chimpanzee (*Pan troglodytes*). *Mol. Gen. Genet.*, **237**, 460-462.

Lundblad, V. and Blackburn, E.H. (1993) An alternative pathway for yeast telomere maintenance rescues EST1-senescence. *Cell*, **73**, 347-360.

Lundblad, V. and Szostak, J.W. (1989) A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell*, **57**, 633-643.

Lundblad, V. and Wright, W.E. (1996) Telomeres and telomerase: a simple picture becomes complex. *Cell*, **87**, 369-375.

Lustig, A.J. and Petes, T.D. (1986) Identification of yeast mutants with altered telomere structure. *Proc. Natl. Acad. Sci. USA*, **83**, 1398-1402.

Macina, R.A., Negorev, D.G., Spais, C., Ruthig, L.A., Hu, X.L. and Riethman, H.C. (1994) Sequence organization of the human chromosome 2q telomere. *Hum. Mol. Genet.*, **3**, 1847-1853.

Makarov, V.L., Hirose, Y. and Langmore, J.P. (1997) Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. *Cell*, **88**, 657-666.

Makarov, V.L., Lejnine, S., Bedoyan, J. and Langmore, J.P. (1993) Nucleosomal organization of telomere-specific chromatin in rat. *Cell*, **73**, 775-787.

Manuelidis, L. and Borden, J. (1988) Reproducible compartmentalization of individual chromosome domains in human CNS cells revealed by *in situ* hybridization and three-dimensional reconstruction. *Chromosoma*, **96**, 397-410.

Marcand, S., Gilson, E. and Shore, D. (1997) A protein-counting mechanism for telomere length regulation in yeast. *Science*, **275**, 986-990.

Martin, G.M., Austad, S.N. and Johnson, T.E. (1996) Genetic analysis of aging - role of oxidative damage and environmental stresses. *Nat. Genet.*, **13**, 25-34.

Mason, J.M. and Biessmann, H. (1995) The unusual telomeres of *Drosophila*. *Trends Genet.*, **11**, 58-62.

Matsumoto, T., Fukui, K., Niwa, O., Sugawara, N., Szostak, J.W. and Yanagida, M. (1987) Identification of healed terminal DNA fragments in linear minichromosomes of *Schizosaccharomyces pombe*. *Mol. Cell. Biol.*, **7**, 4424-4430.

May, C.A., Jeffreys, A.J. and Armour, J.A. (1996) Mutation rate heterogeneity and the generation of allele diversity at the human minisatellite MS205 (D16S309). *Hum. Mol. Genet.*, **5**, 1823-1833.

McClintock, B. (1941) The stability of broken ends of chromosomes in *Zea Mays*. *Genetics*, **26**, 234-282.

McCormick, J.J. and Maher, V.M. (1988) Towards an understanding of the malignant transformation of diploid human fibroblasts. *Mutat. Res.*, **199**, 273-291.

McEachern, M.J. and Blackburn, E.H. (1996) Cap-prevented recombination between terminal telomeric repeat arrays (telomere CPR) maintains telomeres in *Kluyveromyces lactis* lacking telomerase. *Genes Dev.*, **10**, 1822-1834.

McKeown, C., Read, A.P., Dodge, A., Stecko, O., Mercer, A. and Harris, R. (1987) Wolf-Hirschhorn locus is distal to D4S10 on short arm of chromosome 4. *J. Med. Genet.*, **24**, 410-412.

Melek, M. and Shippen, D.E. (1996) Chromosome healing: spontaneous and programmed *de novo* telomere formation by telomerase. *Bioessays*, **18**, 301-308.

Metcalfe, J.A., Parkhill, J., Campbell, L., Stacey, M., Biggs, P., Byrd, P.J. and Taylor, A.M.R. (1996) Accelerated telomere shortening in ataxia telangiectasia. *Nat. Genetics*, **13**, 350-353.

Meyerson, M., Counter, C.M., Eaton, E.N., Ellisen, L.W., Steiner, P., Caddle, S.D., Ziaugra, L., Beijersbergen, R.L., Davidoff, M.J., Liu, Q.Y., Bacchetti, S., Haber, D.A. and Weinberg, R.A. (1997) hEST2, the putative human telomerase catalytic subunit gene, is up- regulated in tumor cells and during immortalization. *Cell*, ***90***, 785-795.

Meyne, J., Ratliff, R.L. and Moyzis, R.K. (1989) Conservation of the human telomere sequence (TTAGGG)n among vertebrates. *Proc. Natl. Acad. Sci. USA*, **86**, 7049-7053.

Mishmar, D., Rahat, A., Scherer, S.W., Nyakatura, G., Hinzmann, B., Kohwi, Y., Mandel-Gutfroind, Y., Lee, J.R., Drescher, B., Sas, D.E., Margalit, H., Platzer, M., Weiss, A., Tsui, L.C., Rosenthal, A. and Kerem, B. (1998) Molecular characterization of a common fragile site (FRA7H) on human chromosome 7 by the cloning of a simian virus 40 integration site. *Proc. Natl. Acad. Sci. USA*, **95**, 8141-8146.

Morales, C.P., Holt, S.E., Ouellette, M., Kaur, K.J., Yan, Y., Wilson, K.S., White, M.A., Wright, W.E. and Shay, J.W. (1999) Absence of cancer-associated changes in human fibroblasts immortalized with telomerase. *Nat. Genet.*, **21**, 115-118.

Morgan, W.F. and Crossen, P.E. (1977) The frequency and distribution of sister chromatid exchanges in human chromosomes. *Hum. Genet.*, **38**, 271-278

Morin, G.B. (1989) The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell*, **59**, 521-529.

Morin, G.B. (1991) Recognition of a chromosome truncation site associated with alpha-thalassemia by human telomerase. *Nature*, **353**, 454-456.

Morris, D.K. and Lundblad, V. (1997) Programmed translational frameshifting in a gene required for yeast telomere replication. *Curr. Biol.*, **7**, 969-976.

Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L. and Wu, J.R. (1988) A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci. USA*, **85**, 6622-6626.

Muller, F., Wicky, C., Spicher, A. and Tobler, H. (1991) New telomere formation after developmentally regulated chromosomal breakage during the process of chromatin diminution in *Ascaris lumbricoides*. *Cell*, **67**, 815-822.

Muller, H.J. (1938) The remaking of chromosomes. *The Collecting Net-Woods Hole.*, **13**, 181-198.

Murnane, J.P., Sabatier, L., Marder, B.A. and Morgan, W.F. (1994) Telomere dynamics in an immortal human cell line. *EMBO J.*, **13**, 4953-4962.

Murnane, J.P. and Yu, L.C. (1993) Acquisition of telomere repeat sequences by transfected DNA integrated at the site of a chromosome break. *Mol. Cell. Biol.*, **13**, 977-983.

Murray, A.W. and Szostak, J.W. (1983) Construction of artificial chromosomes in yeast. *Nature*, **305**, 189-193.

Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W.H., Lingner, J., Harley, C.B. and Cech, T.R. (1997) Telomerase catalytic subunit homologs from fission yeast and human. *Science*, **277**, 955-959.

Nakayama, J.I., Saito, M., Nakamura, H., Matsuura, A. and Ishikawa, F. (1997) TLP1: A gene encoding a protein component of mammalian telomerase is a novel member of WD repeats family. *Cell*, **88**, 875-884.

National Institutes of Health and Institute of Molecular Medicine collaboration. (1996) A complete set of human telomeric probes and their clinical application. *Nat. Genet.*, **14**, 86-89.

NIH/CEPH Collaborative Mapping Group. (1992) A comprehensive genetic linkage map of the human genome. *Science*, **258**, 67-86.

Nelson, W.G. and Kastan, M.B. (1994) DNA strand breaks: the DNA template alterations that trigger p53- dependent DNA damage response pathways. *Mol. Cell. Biol.*, **14**, 1815-1823.

Nesslinger, N.J., Gorski, J.L., Kurczynski, T.W., Shapira, S.K., Siegelbartelt, J., Dumanski, J.P., Cullen, R.F., French, B.N. and McDermid, H.E. (1994) Clinical, cytogenetic, and molecular characterization of seven patients with deletions of chromosome 22q13.3. *Am. J. Hum. Genet.*, **54**, 464-472.

Newbold, R.F. (1997) Genetic control of telomerase and replicative senescence in human and rodent cells. *Ciba Found. Symp.*, **211**, 177-189.

Nicholls, R.D., Fischel-Ghodsian, N. and Higgs, D.R. (1987) Recombination at the human alpha-globin gene cluster: sequence features and topological constraints. *Cell*, **49**, 369-378.

Niida, H., Matsumoto, T., Satoh, H., Shiwa, M., Tokutake, Y., Furuichi, Y. and Shinkai, Y. (1998) Severe growth defect in mouse cells lacking the telomerase RNA component. *Nat. Genet.*, **19**, 203-206.

Nimmo, E.R., Cranston, G. and Allshire, R.C. (1994) Telomere-associated chromosome breakage in fission yeast results in variegated expression of adjacent genes. *EMBO J.*, **13**, 3801-3811.

Nowell, P.C. (1986) Mechanisms of tumor progression. *Cancer Res.*, 46, 2203-2207.

Nugent, C.I., Hughes, T.R., Lue, N.F. and Lundblad, V. (1996) Cdc13p - a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance. *Science*, **274**, 249-252.

Oberle, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M.F. and Mandel, J.L. (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science*, **252**, 1097-1102.

Okazaki, S., Tsuchida, K., Maekawa, H., Ishikawa, H. and Fujiwara, H. (1993) Identification of a pentanucleotide telomeric sequence, (TTAGG)n, in the silkworm *Bombyx mori* and in other insects. *Mol. Cell. Biol.*, **13**, 1424-1432.

Olovnikov, A.M. (1971) Principle of marginotomy in template synthesis of polynucleotides. *Dokl. Akad. Nauk SSSR*, **201**, 1496-1499.

Olovnikov, A.M. (1973) A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *J. Theor. Biol.*, **41**, 181-190.

Palladino, F., Laroche, T., Gilson, E., Axelrod, A., Pillus, L. and Gasser, S.M. (1993) Sir3 and Sir4 proteins are required for the positioning and integrity of yeast telomeres. *Cell*, **75**, 543-555.

Pandita, T.K., Pathak, S. and Geard, C.R. (1995) Chromosome end associations, telomeres and telomerase activity in ataxia-telangiectasia cells. *Cytogenet. Cell Genet.*, **71**, 86-93.

Pardue, M.L., Danilevskaya, O.N., Lowenhaupt, K., Slot, F. and Traverse, K.L. (1996) *Drosophila* telomeres: new views on chromosome evolution. *Trends Genet.*, **12**, 48-52.

Pereira-Smith, O.M. and Smith, J.R. (1988) Genetic analysis of indefinite division in human cells: identification of four complementation groups. *Proc. Natl. Acad. Sci. USA*, **85**, 6042-6046.

Petracek, M.E., Lefebvre, P.A., Silflow, C.D. and Berman, J. (1990) *Chlamydomonas* telomere sequences are A+T-rich but contain three consecutive G-C base pairs. *Proc. Natl. Acad. Sci. USA*, **87**, 8222-8226.

Petracek, M.E., Konkel, L.M., Kable, M.L. and Berman, J. (1994) A *Chlamydomonas* protein that binds single-stranded G-strand telomere DNA. *EMBO J.*, **13**, 3648-3658.

Pierce, G.B. and Speers, W.C. (1988) Tumors as caricatures of the process of tissue renewal: prospects for therapy by directing differentiation. *Cancer Res.*, **48**, 1996-2004.

Pieretti, M., Zhang, F.P., Fu, Y.H., Warren, S.T., Oostra, B.A., Caskey, C.T. and Nelson, D.L. (1991) Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell*, **66**, 817-822.

Ponten, J. (1976) The relationship between *in vitro* transformation and tumor formation *in vivo*. *Biochim. Biophys. Acta*, **458**, 397-422.

Potten, C.S. and Loeffler, M. (1990) Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt. *Development*, **110**, 1001-1020.

Prescott, D.M. (1994) The DNA of ciliated protozoa. *Microbiol. Rev.*, 58, 233-267.

Prescott, J. and Blackburn, E.H. (1997) Functionally interacting telomerase RNAs in the yeast telomerase complex. *Genes Dev.*, **11**, 2790-2800.

Price, C.M. (1990) Telomere structure in *Euplotes crassus* - characterization of DNA-protein interactions and isolation of a telomere-binding protein. *Mol. Cell. Biol.*, **10**, 3421-3431.

Proudfoot, N.J., Gil, A. and Maniatis, T. (1982) The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. *Cell*, **31**, 553-563.

Prowse, K.R., Avilion, A.A. and Greider, C.W. (1993) Identification of a nonprocessive telomerase activity from mouse cells. *Proc. Natl. Acad. Sci. USA*, **90**, 1493-1497.

Prowse, K.R. and Greider, C.W. (1995) Developmental and tissue-specific regulation of mouse telomerase and telomere length. *Proc. Natl. Acad. Sci. USA*, **92**, 4818-4822.

Quarrell, O.W., Snell, R.G., Curtis, M.A., Roberts, S.H., Harper, P.S. and Shaw, D.J. (1991) Paternal origin of the chromosomal deletion resulting in Wolf- Hirschhorn syndrome. *J. Med. Genet.*, **28**, 256-259.

Ramakrishnan, S., Sharma, H.W., Farris, A.D., Kaufman, K.M., Harley, J.B., Collins, K., Pruijn, G., vanVenrooij, W.J., Martin, M.L. and Narayanan, R. (1997) Characterization of human telomerase complex. *Proc. Natl. Acad. Sci. USA*, **94**, 10075-10079.

Rasmussen, S.W. and Holm, P.B. (1980) Mechanics of meiosis. *Hereditas*, **93**, 187-216.

Ray, A. and Runge, K.W. (1999) The yeast telomere length counting machinery is sensitive to sequences at the telomere-nontelomere junction. *Mol. Cell. Biol.*, **19**, 31-45.

Reiner, O., Carrozzo, R., Shen, Y., Wehnert, M., Faustinella, F., Dobyns, W.B., Caskey, C.T. and Ledbetter, D.H. (1993) Isolation of a Miller-Dieker lissencephaly gene containing G protein beta-subunit-like repeats. *Nature*, **364**, 717-721.

Renauld, H., Aparicio, O.M., Zierath, P.D., Billington, B.L., Chhablani, S.K. and Gottschling, D.E. (1993) Silent domains are assembled continuously from the telomere and are defined by promoter distance and strength, and by SIR3 dosage. *Genes Dev.*, **7**, 1133-1145.

Richards, E.J. and Ausubel, F.M. (1988) Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell*, **53**, 127-136.

Riethman, H.C., Moyzis, R.K., Meyne, J., Burke, D.T. and Olson, M.V. (1989) Cloning human telomeric DNA fragments into *Saccharomyces cerevisiae* using a yeast artificial chromosome vector. *Proc. Natl. Acad. Sci. USA*, **86**, 6240-6244.

Rogan, E.M., Bryan, T.M., Hukku, B., Maclean, K., Chang, A.C., Moy, E.L., Englezou, A., Warneford, S.G., Dalla-Pozza, L. and Reddel, R.R. (1995) Alterations in p53 and $p16^{INK4}$ expression and telomere length during spontaneous immortalization of Li-Fraumeni syndrome fibroblasts. *Mol. Cell. Biol.*, **15**, 4745-4753.

Rogers, J. (1985) Oncogene chromosome breakpoints and *Alu* sequences. *Nature*, **317**, 559.

Romero, D.P. and Blackburn, E.H. (1991) A conserved secondary structure for telomerase RNA. *Cell*, **67**, 343-353.

Rosenberg, M., Hui, L., Ma, J.L., Nusbaum, H.C., Clark, K., Robinson, L., Dziadzio, L., Swain, P.M., Keith, T., Hudson, T.J., Biesecker, L.G. and Flint, J. (1997) Characterization of short tandem repeats from thirty-one human telomeres. *Genome Res.*, **7**, 917-923.

Roth, D.B., Chang, X.B. and Wilson, J.H. (1989) Comparison of filler DNA at immune, nonimmune, and oncogenic rearrangements suggests multiple mechanisms of formation. *Mol. Cell. Biol.*, **9**, 3049-3057.

Rouyer, F., de la Chapelle, A., Andersson, M. and Weissenbach, J. (1990) An interspersed repeated sequence specific for human subtelomeric regions. *EMBO J.*, **9**, 505-514.

Royle, N.J. (1995) The proterminal regions and telomeres of human chromosomes. *Adv. Genet.*, **32**, 273-315.

Royle, N.J., Baird, D.M. and Jeffreys, A.J. (1994) A subterminal satellite located adjacent to telomeres in chimpanzees is absent from the human genome. *Nat. Genet.*, **6**, 52-56.

Royle, N.J., Clarkson, R.E., Wong, Z. and Jeffreys, A.J. (1988) Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics*, **3**, 352-360.

Royle, N.J., Hill, M.C. and Jeffreys, A.J. (1992) Isolation of telomere junction fragments by anchored polymerase chain reaction. *Proc. R. Soc. Lond. B-*, **247**, 57-61.

Runge, K.W. and Zakian, V.A. (1996) TEL2, an essential gene required for telomere length regulation and telomere position effect in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **16**, 3094-3105.

Saccone, S., De Sario, A., Della Valle, G. and Bernardi, G. (1992) The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes. *Proc. Natl. Acad. Sci. USA*, **89**, 4913-4917.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487-491.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Second edition. Cold Spring Harbour Press.

Sandell, L.L. and Zakian, V.A. (1993) Loss of a yeast telomere - arrest, recovery, and chromosome loss. *Cell*, **75**, 729-739.

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

Schinzel, A., Auf der Maur, P. and Moser, H. (1977) Partial deletion of long arm of chromosome 11[del(11)(q23)]: Jacobsen syndrome. Two new cases and review of the clinical findings. *J. Med. Genet.*, **14**, 438-444.

Schmid, C.W. and Jelinek, W.R. (1982) The *Alu* family of dispersed repetitive sequences. *Science*, **216**, 1065-1070.

Schneider, S.S., Schick, C., Fish, K.E., Miller, E., Pena, J.C., Treter, S.D., Hui, S.M. and Silverman, G.A. (1995) A serine proteinase inhibitor locus at 18q21.3 contains a tandem duplication of the human squamous cell carcinoma antigen gene. *Proc. Natl. Acad. Sci. USA*, **92**, 3147-3151.

Shampay, J., Szostak, J.W. and Blackburn, E.H. (1984) DNA sequences of telomeres maintained in yeast. *Nature*, **310**, 154-157.

Shay, J.W., Pereira-Smith, O.M. and Wright, W.E. (1991a) A role for both RB and p53 in the regulation of human cellular senescence. *Exp. Cell Res.*, **196**, 33-39.

Shay, J.W. and Wright, W.E. (1989) Quantitation of the frequency of immortalization of normal human diploid fibroblasts by SV40 large T-antigen. *Exp. Cell Res.*, **184**, 109-118.

Shay, J.W. and Wright, W.E. (1996) The reactivation of telomerase activity in cancer progression. *Trends Genet.*, **12**, 129-131.

Shay, J.W., Wright, W.E., Brasiskyte, D. and Vanderhaegen, B.A. (1993) E6 of human papillomavirus type-16 can overcome the M1 stage of immortalization in human mammary epithelial cells but not in human fibroblasts. *Oncogene*, **8**, 1407-1413.

Shay, J.W., Wright, W.E. and Werbin, H. (1991b) Defining the molecular mechanisms of human cell immortalization. *Biochim. Biophys. Acta.*, **1072**, 1-7.

Sheen, F.M. and Levis, R.W. (1994) Transposition of the LINE-like retrotransposon TART to *Drosophila* chromosome termini. *Proc. Natl. Acad. Sci. USA*, **91**, 12510-12514.

Sheng, H., Hou, Z., Schierer, T., Dobbs, D.L. and Henderson, E. (1995) Identification and characterization of a putative telomere end-binding protein from *Tetrahymena thermophila*. *Mol. Cell. Biol.*, **15**, 1144-1153.

Shiloh, Y. (1995) Ataxia telangiectasia - closer to unraveling the mystery. *Eur. J. Hum. Genet.*, **3**, 116-138.

Shippenlentz, D. and Blackburn, E.H. (1990) Functional evidence for an RNA template in telomerase. *Science*, **247**, 546-552.

Shore, D. (1997) Telomeres - different means to common ends. *Nature*, **385**, 676-677.

Shore, D. and Nasmyth, K. (1987) Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. *Cell*, **51**, 721-732.

Singer, M.S. and Gottschling, D.E. (1994) TLC1 - template RNA component of *Saccharomyces cerevisiae* telomerase. *Science*, **266**, 404-409.

Slagboom, P.E., Droog, S. and Boomsma, D.I. (1994) Genetic determination of telomere size in humans: a twin study of three age groups. *Am. J. Hum. Genet.*, **55**, 876-882.

Slagboom, P.E. and Vijg, J. (1992) The dynamics of genome organization and expression during the aging process. *Ann. N. Y. Acad. Sci.*, **673**, 58-69.

Smit, A.F. and Riggs, A.D. (1995) MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucleic Acids Res.*, **23**, 98-102.

Smith, S. and de Lange, T. (1997) TRF1, a mammalian telomeric protein. *Trends Genet.*, **13**, 21-26.

Smith, S., Giriat, I., Schmitt, A. and de Lange, T. (1998) Tankyrase, a poly(ADP-ribose) polymerase at human telomeres. *Science*, **282**, 1484-1487.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.*, **98**, 503-517.

Spangler, E.A., Ryan, T. and Blackburn, E.H. (1988) Developmentally regulated telomere addition in *Tetrahymena thermophila*. *Nucleic Acids Res.*, **16**, 5569-5585.

Sprung, C.N., Sabatier, L. and Murnane, J.P. (1996) Effect of telomere length on telomeric gene-expression. *Nucleic Acids Res.*, **24**, 4336-4340.

Starling, J.A., Maule, J., Hastie, N.D. and Allshire, R.C. (1990) Extensive telomere repeat arrays in mouse are hypervariable. *Nucleic Acids Res.*, **18**, 6881-6888.

Stein, G.H. (1985) SV40-transformed human fibroblasts: evidence for cellular aging in pre-crisis cells. *J. Cell. Physiol.*, **125**, 36-44.

Steiner, B.R., Hidaka, K. and Futcher, B. (1996) Association of the Est1 protein with telomerase activity in yeast. *Proc. Natl. Acad. Sci. USA*, **93**, 2817-2821.

Sundquist, W.I. and Klug, A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825-829.

Sutherland, G.R. (1982) Heritable fragile sites on human chromosomes. VIII. Preliminary population cytogenetic data on the folic-acid-sensitive fragile sites. *Am. J. Hum. Genet.*, **34**, 452-458.

Sutherland, G.R. and Richards, R.I. (1995) The molecular basis of fragile sites in human chromosomes. *Curr. Opin. Genet. Dev.*, **5**, 323-327.

Therman, E., Trunca, C., Kuhn, E.M. and Sarto, G.E. (1986) Dicentric chromosomes and the inactivation of the centromere. *Hum. Genet.*, **72**, 191-195

Thies, U., Back, E., Wolff, G., Schroeder-Kurth, T., Hager, H.D. and Schroder, K. (1992) Clinical, cytogenetic and molecular investigations in three patients with Wolf-Hirschhorn syndrome. *Clin. Genet.*, **42**, 201-205.

Thoraval, D., Asakawa, J., Kodaira, M., Chang, C., Radany, E., Kuick, R., Lamb, B., Richardson, B., Neel, J.V., Glover, T. and Hanash, S. (1996) A methylated human 9-kb repetitive sequence on acrocentric chromosomes is homologous to a subtelomeric repeat in chimpanzees. *Proc. Natl. Acad. Sci. USA*, **93**, 4442-4447.

Tommerup, H., Dousmanis, A. and Delange, T. (1994) Unusual chromatin in human telomeres. *Mol. Cell. Biol.*, **14**, 5777-5785.

Van der Haegen, B.A. and Shay, J.W. (1993) Immortalization of human mammary epithelial cells by SV40 large T- antigen involves a two step mechanism. *In Vitro Cell. Dev. Biol.*, **29A**, 180-182.

Van der Ploeg, L.H., Liu, A.Y. and Borst, P. (1984) Structure of the growing telomeres of Trypanosomes. *Cell*, **36**, 459-468.

van Deutekom, J.C., Lemmers, R.J., Grewal, P.K., van Geel, M., Romberg, S., Dauwerse, H.G., Wright, T.J., Padberg, G.W., Hofker, M.H., Hewitt, J.E. and Frants, R.R. (1996) Identification of the first gene (FRG1) from the FSHD region on human chromosome 4q35. *Hum. Mol. Genet.*, **5**, 581-590.

van Deutekom, J.C., Wijmenga, C., van Tienhoven, E.A., Gruter, A.M., Hewitt, J.E., Padberg, G.W., van Ommen, G.J., Hofker, M.H. and Frants, R.R. (1993) FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum. Mol. Genet.*, **2**, 2037-2042.

van Steensel, B. and de Lange, T. (1997) Control of telomere length by the human telomeric protein TRF1. *Nature*, **385**, 740-743.

Vanin, E.F., Henthorn, P.S., Kioussis, D., Grosveld, F. and Smithies, O. (1983) Unexpected relationships between four large deletions in the human beta-globin gene cluster. *Cell*, **35**, 701-709.

Vaziri, H. and Benchimol, S. (1998) Reconstitution of telomerase activity in normal human cells leads to elongation of telomeres and extended replicative life span. *Curr. Biol.*, **8**, 279-282.

Vaziri, H., Dragowska, W., Allsopp, R.C., Thomas, T.E., Harley, C.B. and Lansdorp, P.M. (1994) Evidence for a mitotic clock in human hematopoietic stem cells: loss of telomeric DNA with age. *Proc. Natl. Acad. Sci. USA*, **91**, 9857-9860.

Vaziri, H., Schachter, F., Uchida, I., Wei, L., Zhu, X.M., Effros, R., Cohen, D. and Harley, C.B. (1993) Loss of telomeric DNA during aging of normal and trisomy-21 human lymphocytes. *Am. J. Hum. Genet.*, **52**, 661-667.

Vergnaud, G., Mariat, D., Zoroastro, M. and Lauthier, V. (1991) Detection of single and multiple polymorphic loci by synthetic tandem repeats of short oligonucleotides. *Electrophoresis*, **12**, 134-140.

Vickers, M.A., Vyas, P., Harris, P.C., Simmons, D.L. and Higgs, D.R. (1993) Structure of the human 3-methyladenine DNA glycosylase gene and localization close to the 16p telomere. *Proc. Natl. Acad. Sci. USA*, **90**, 3437-3441.

Vincent, A., Heitz, D., Petit, C., Kretz, C., Oberle, I. and Mandel, J.L. (1991) Abnormal pattern detected in fragile-X patients by pulsed-field gel electrophoresis. *Nature*, **349**, 624-626.

Virta-Pearlman, V., Morris, D.K. and Lundblad, V. (1996) Est1 has the properties of a single-stranded telomere end-binding protein. *Genes Dev.*, **10**, 3094-3104.

Vocero-Akbani, A., Helms, C., Wang, J.C., Sanjurjo, F.J., Korte-Sarfaty, J., Veile, R.A., Liu, L., Jauch, A., Burgess, A.K., Hing, A.V., Holt, M.S., Ramachandra, S., Whelan, A.J., Anker, R., Ahrent, L., Chen, M., Gavin, M.R., Iannantuoni, K., Morton, S.M., Pandit, S.D., Read, C.M., Steinbrueck, T., Warlick, C., Smoller, D.A. and Donis-Keller, H. (1996) Mapping human telomere regions with YAC and P1 clones: chromosome-specific markers for 27 telomeres including 149 STSs and 24 polymorphisms for 14 proterminal regions. *Genomics*, **36**, 492-506.

Voullaire, L.E., Webb, G.C. and Leversha, M.A. (1987) Chromosome deletion at 11q23 in an abnormal child from a family with inherited fragility at 11q23. *Hum. Genet.*, **76**, 202-204.

Vourc'h, C., Taruscio, D., Boyle, A.L. and Ward, D.C. (1993) Cell cycle-dependent distribution of telomeres, centromeres, and chromosome-specific subsatellite domains in the interphase nucleus of mouse lymphocytes. *Exp. Cell Res.*, **205**, 142-151.

Vyas, P., Vickers, M.A., Simmons, D.L., Ayyub, H., Craddock, C.F. and Higgs, D.R. (1992) Cis-acting sequences regulating expression of the human alpha-globin cluster lie within constitutively open chromatin. *Cell*, **69**, 781-793.

Wallace, B.M. and Hulten, M.A. (1985) Meiotic chromosome pairing in the normal human female. *Ann. Hum. Genet.*, **49**, 215-226.

Wang, H. and Blackburn, E.H. (1997) *De novo* telomere addition by *Tetrahymena* telomerase in vitro. *EMBO J.*, **16**, 866-879.

Wang, S.S. and Zakian, V.A. (1990) Telomere-telomere recombination provides an express pathway for telomere acquisition. *Nature*, **345**, 456-458.

Wareham, K.A., Lyon, M.F., Glenister, P.H. and Williams, E.D. (1987) Age-related reactivation of an X-linked gene. *Nature*, **327**, 725-727.

Watson, J.D. (1972) Origin of concatemeric T7 DNA. *Nat. New Biol.*, **239**, 197-201.

Weber, B., Collins, C., Robbins, C., Magenis, R.E., Delaney, A.D., Gray, J.W. and Hayden, M.R. (1990) Characterization and organization of DNA sequences adjacent to the human telomere associated repeat (TTAGGG)n. *Nucleic Acids Res.*, **18**, 3353-3361.

Weber, S., Schmid, M., Meyer, J., Cooke, H.J. and Lipps, H.J. (1993) A linear vector carrying human telomeres is replicated in unfertilized eggs of *Xenopus laevis*. *Cell Biol. Int.*, **17**, 623-624.

Weinert, T.A. and Hartwell, L.H. (1993) Cell cycle arrest of cdc mutants and specificity of the RAD9 checkpoint. *Genetics*, **134**, 63-80.

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

Wellinger, R.J., Ethier, K., Labrecque, P. and Zakian, V.A. (1996) Evidence for a new step in telomere maintenance. *Cell*, **85**, 423-433.

Wellinger, R.J., Wolf, A.J. and Zakian, V.A. (1992) Use of non-denaturing Southern hybridization and two dimensional agarose gels to detect putative intermediates in telomere replication in *Saccharomyces cerevisiae*. *Chromosoma*, **102**, S150-156.

Wellinger, R.J., Wolf, A.J. and Zakian, V.A. (1993) Origin activation and formation of single-strand TG_{1-3} tails occur sequentially in late S phase on a yeast linear plasmid. *Mol. Cell. Biol.*, 13, 4057-4065.

Wells, R.A., Germino, G.G., Krishna, S., Buckle, V.J. and Reeders, S.T. (1990) Telomere-related sequences at interstitial sites in the human genome. *Genomics*, **8**, 699-704.

Weng, N.P., Granger, L. and Hodes, R.J. (1997) Telomere lengthening and telomerase activation during human B cell differentiation. *Proc. Natl. Acad. Sci. USA*, **94**, 10827-10832.

Wijmenga, C., Hewitt, J.E., Sandkuijl, L.A., Clark, L.N., Wright, T.J., Dauwerse, H.G., Gruter, A.M., Hofker, M.H., Moerer, P., Williamson, R. and et al. (1992) Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat. Genet.*, **2**, 26-30

Wilkie, A.O., Buckle, V.J., Harris, P.C., Lamb, J., Barton, N.J., Reeders, S.T., Lindenbaum, R.H., Nicholls, R.D., Barrow, M., Bethlenfalvay, N.C. and *et al.* (1990) Clinical features and molecular analysis of the alpha-thalassemia/mental retardation syndromes. I. Cases due to deletions involving chromosome band 16p13.3. *Am. J. Hum. Genet.*, **46**, 1112-1126.

Wilkie, A.O., Higgs, D.R., Rack, K.A., Buckle, V.J., Spurr, N.K., Fischel-Ghodsian, N., Ceccherini, I., Brown, W.R. and Harris, P.C. (1991) Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome **16**. *Cell*, 64, 595-606.

Williamson, J.R., Raghuraman, M.K. and Cech, T.R. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, **59**, 871-880.

Wilson, L.C., Leverton, K., Oude Luttikhuis, M.E., Oley, C.A., Flint, J., Wolstenholme, J., Duckett, D.P., Barrow, M.A., Leonard, J.V., Read, A.P. and et al. (1995) Brachydactyly and mental retardation: an Albright hereditary osteodystrophy-like syndrome localized to 2q37. *Am. J. Hum. Genet.*, **56**, 400-407.

Wong, A.C.C., Ning, Y., Flint, J., Clark, K., Dumanski, J.P., Ledbetter, D.H. and McDermid, H.E. (1997) Molecular characterization of a 130-kb terminal microdeletion at 22q in a child with mild mental retardation. *Am. J. Hum. Genet.*, **60**, 113-120.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A.J. (1987) Characterization of a panel of highly variable minisatellites cloned from human DNA. *Ann. Hum. Genet.*, **51**, 269-288.

Wotton, D. and Shore, D. (1997) A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in *Saccharomyces cerevisiae*. *Genes Dev.*, **11**, 748-760.

Wright, W.E., Brasiskyte, D., Piatyszek, M.A. and Shay, J.W. (1996a) Experimental elongation of telomeres extends the lifespan of immortal cell x normal cell Hybrids. *EMBO J.*, **15**, 1734-1741.

Wright, T.J., Clemens, M., Quarrell, O. and Altherr, M.R. (1998) Wolf-Hirschhorn and Pitt-Rogers-Danks syndromes caused by overlapping 4p deletions. *Am. J. Med. Genet.*, **75**, 345-350.

Wright, J.H., Gottschling, D.E. and Zakian, V.A. (1992) *Saccharomyces* telomeres assume a non-nucleosomal chromatin structure. *Genes Dev.*, **6**, 197-210.

Wright, W.E., Pereira-Smith, O.M. and Shay, J.W. (1989) Reversible cellular senescence: implications for immortalization of normal human diploid fibroblasts. *Mol. Cell. Biol.*, **9**, 3088-3092.

Wright, W.E., Piatyszek, M.A., Rainey, W.E., Byrd, W. and Shay, J.W. (1996b) Telomerase activity in human germline and embryonic tissues and cells. *Dev. Genet.*, **18**, 173-179.

Wright, W.E. and Shay, J.W. (1995) Time, telomeres and tumors - is cellular senescence more than an anticancer mechanism. *Trends Cell Biol.*, **5**, 293-297.

Wright, J.H. and Zakian, V.A. (1995) Protein-DNA interactions in soluble telosomes from *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **23**, 1454-1460.

Yao, M.C., Zheng, K. and Yao, C.H. (1987) A conserved nucleotide sequence at the sites of developmentally regulated chromosomal breakage in *Tetrahymena*. *Cell*, **48**, 779-788.

Yasumoto, S., Kunimura, C., Kikuchi, K., Tahara, H., Ohji, H., Yamamoto, H., Ide, T. and Utakoji, T. (1996) Telomerase activity in normal human epithelial cells. *Oncogene*, **13**, 433-439.

Youngren, K., Jeanclos, E., Aviv, H., Kimura, M., Stock, J., Hanna, M., Skurnick, J., Bardeguez, A. and Aviv, A. (1998) Synchrony in telomere length of the human fetus. *Hum. Genet.*, **102**, 640-643.

Yu, G.L. and Blackburn, E.H. (1991) Developmentally programmed healing of chromosomes by telomerase in *Tetrahymena*. *Cell*, **67**, 823-832.

Yu, G.L., Bradley, J.D., Attardi, L.D. and Blackburn, E.H. (1990) In vivo alteration of telomere sequences and senescence caused by mutated *Tetrahymena* telomerase RNAs. *Nature*, **344**, 126-132.

Yuasa, T., Yoshiki, T., Tanaka, T., Kim, C.J., Isono, T. and Okada, Y. (1998) Expression of uroplakin Ib and uroplakin III genes in tissues and peripheral blood of patients with transitional cell carcinoma. *Jpn. J. Cancer Res.*, **89**, 879-882.

Zakian, V.A. (1989) Structure and function of telomeres. *Annu. Rev. Genet.*, **23**, 579-604.

Zakian, V.A. (1995) Telomeres: beginning to understand the end. *Science*, **270**, 1601-1607.

Zakian, V.A. (1996) Structure, function, and replication of *Saccharomyces cerevisiae* telomeres. *Annu. Rev. Genet.*, **30**, 141-172.

Zakian, V.A. (1997) Life and cancer without telomerase. *Cell*, **91**, 1-3.

Zakian, V.A. and Blanton, H.M. (1988) Distribution of telomere-associated sequences on natural chromosomes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **8**, 2257-2260.

Zakian, V.A., Runge, K. and Wang, S.S. (1990) How does the end begin? Formation and maintenance of telomeres in ciliates and yeast. *Trends Genet.*, **6**, 12-16.

Zhong, Z., Shiue, L., Kaplan, S. and de Lange, T. (1992) A mammalian factor that binds telomeric TTAGGG repeats in vitro. *Mol. Cell. Biol.*, **12**, 4834-4843.