MOLECULAR GENETIC ANALYSIS OF HUMAN 8p INVERSION DUPLICATION CHROMOSOMES

by

TANYA N. NELSON

B.Sc. The University of British Columbia, 1993

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Medical Genetics Programme)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September 1998

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of MEDICAL GENETICS

The University of British Columbia
Vancouver, Canada

Date SEPTEMBER 28, 1998

Abstract

Inversion duplications of the short arm of chromosome 8 (8p) with common morphology have been described in over 60 mentally retarded individuals. These aberrant chromosomes contain material derived from both maternal chromosomes, separated by a single copy region at a common center of symmetry, with deletion of distal subtelomeric sequences. A mechanism mediated by inverted repetitive elements may explain the recurrence of these similar aberrant chromosomes in unrelated patients. A megasatellite repeated sequence, localized to chromosome 8p23, was investigated as a candidate for the proposed repetitive elements. Cosmid clones isolated from a single chromosome 8 library contained three classes of megasatellite. Megasatellite-containing YAC clones map to two locations on chromosome 8p consistent with the flanking positions predicted by the known extent of the single copy region. Although the megasatellite sequences span 16 kb, all cosmids containing the megasatellite cross hybridize outside of this region. A BAC clone, that lacks the megasatellite but contains these cross-hybridizing sequences within which the megasatellite is embedded, was used as a FISH probe to metaphase chromosomes. Hybridization occurred at multiple locations throughout the genome, including 8p23. Therefore, the 8p copies of the megasatellite are each embedded within a region of at least 160 kb that is itself reiterated throughout the genome. These results suggest that the megasatellite, embedded within a large reiterated region of the genome, may be involved in the generation of inversion duplication (8p) chromosomes by providing a site for anomalous interchromosomal recombination. Investigation of a patient where the single copy region could not be detected suggests that, in rare cases, other mechanisms may be involved.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to express my gratitude to my research supervisors, Drs. Fred Dill and Stephen Wood, for their support, their guidance, and for the opportunity to learn from them. To my supervisory committee, Drs. Sylvie Langlois, Wendy Robinson, and Muriel Harris, thank you for your encouragement, insight, and for always being available. Special thanks to Mike Schertzer for his knowledge and unflagging patience. Thank you to Stanya Jurenka for sharing her clinical knowledge and for all of her kind words. Thank you to Dr. Dagmar Kalousek, the members of her laboratory, and the members of the diagnostic cytogenetic laboratories at B.C. Children's Hospital and the Royal Columbian Hospital, for their time and assistance.

To my family and friends, thank you for listening and for believing in me. And finally, thank you to Colin O'Connor for always reminding me of what is most important.

Chapter 1: Introduction

1.1 Human Chromosome Aberrations

Chromosome abnormalities are seen in approximately 1% of live-born children (Jacobs, 1990). These aberrations may be numerical, resulting from the gain or loss of an entire chromosome, or may be structural, involving rearrangement of single or multiple chromosomes.

The gain or loss of a chromosome is generally believed to occur by the process of nondisjunction, the failure of chromosomes to disjoin. The mechanism leading to this nondisjunction is still under investigation; however, there may be chromosome specific mechanisms involved (Abruzzo and Hassold, 1995). Conversely, structural abnormalities must arise by intrachromosomal or interchromosomal interactions leading to changes in chromosome morphology, rather than a failure to undergo proper disjunction.

1.2 Translocations

The most common *de novo* structural abnormality is a reciprocal translocation (Warburton 1984), a balanced rearrangement involving exchange of material between two non-homologous chromosomes. The mechanism by which reciprocal translocations arise is unclear. However, analysis of translocations ascertained through fetal wastage indicates that the location of the chromosome breakpoints appears to be random (Boué *et al.*, 1985). In some cases of X;autosome reciprocal translocation, a few base pairs of homology are found at the sites of rearrangement, but otherwise there are no remarkable features of the genome at these locations (Bodrug *et al.*, 1987, Giacalone and Francke, 1992). These homologous sequence motifs have

not been found at all sequenced X;autosome translocation breakpoints (Bodrug *et al.*, 1991), therefore, sequence homology may not be required for chromosomal translocation.

A Robertsonian translocation chromosome results from the rearrangement of two acrocentric chromosomes such that the short (p) arms are lost and the long (q) arms join to form the derivative chromosome. The mechanism generating these chromosomes is not known. Ribosomal RNA genes and several families of repetitive (satellite) DNA are present on the p arms of all acrocentric chromosomes (Choo, 1988, Choo, 1990, Trowell, 1993). Most Robertsonian translocation chromosomes are dicentric, with breakpoints in the p arm between the β-satellite and α-satellite DNA (Cheung et al., 1990, Wolff and Schwartz, 1992, Sullivan *et al.*, 1996).). More specifically, these translocations most often involve chromosomes 14 and 21 (14q21q), and chromosomes 13 and 14 (13q14q) (Choo *et al.*, 1988, Therman *et al.*, 1989) with breakpoints occurring between two chromosome 14 specific subfamilies of satellite III sequences, and distal to the satellite I sequences of chromosomes 13 and 21 (Earle *et al.*, 1992, Kalitsis *et al.*, 1993, Han *et al.*, 1994, Sullivan et al., 1996). The breakpoints in other non-homologous Robertsonian translocations are more variable (Page *et al.*, 1996, Sullivan *et al.*, 1996). Because there is a difference in occurrence rate and breakpoint variability between the 'common' non-homologous Robertsonian translocations and 'rarer' non-homologous Robertsonian translocations, it has been suggested that the mechanisms involved in the generation of these chromosomes differ (Page *et al.*, 1996). The proposed mechanisms are based on interaction and recombination at homologous repetitive p arm sequences leading to the formation of Robertsonian chromosomes (Ferguson-Smith, 1967, Therman, 1980, Guichaoua et al., 1986, Choo et al., 1988). The (13q14q) and (14q21q) chromosomes may arise by preferential

interaction of homologous repetitive sequences predicted to be present on chromosome 14 in opposite direction to those on chromosomes 13 and 21 (Choo *et al.*, 1988, Therman *et al.*, 1989, Sullivan *et al.*, 1996). Interactions between other combinations of non-homologous acrocentric chromosomes may occur at smaller regions of homology, where recombination may be less likely to occur, and therefore, less frequently lead to Robertsonian translocations (Page et al., 1996). Although satellite DNA is implicated in the formation of at least some of these chromosomes, it is unclear whether interactions at these repetitive sequences initiate recombination or whether ribosomal RNA interactions are responsible for the initial interaction of the acrocentric chromosomes, followed by recombination at more proximal sequences. (Schmickel and Knoller, 1977, Sullivan et al., 1996).

1.3 Duplications, Deletions, Inversions

Many examples of these types of rearrangements have been reported. However, certain regions of the genome appear to be more likely to undergo rearrangement. In many cases of recurrent rearrangement, such as the duplication in Charcot-Marie-Tooth type 1A (CMT1A) (Lupski *et al.*, 1991), sequences flanking the regions involved in the rearrangement are highly homologous, and evidence of recombination between these repetitive sequences can be shown on the aberrant chromosome. This has led to speculation that the underlying mechanism leading to recurrent rearrangements is similar, and is based on inter- and intra-chromosomal interactions at repetitive sequences. The outcome of these interactions would be dependent on the orientation of the repetitive sequences relative to one another, and on whether the interaction, and subsequent recombination, is between homologues, between sister chromatids, or within a single chromatid (figure 1) (Robinson *et al.*, 1998).

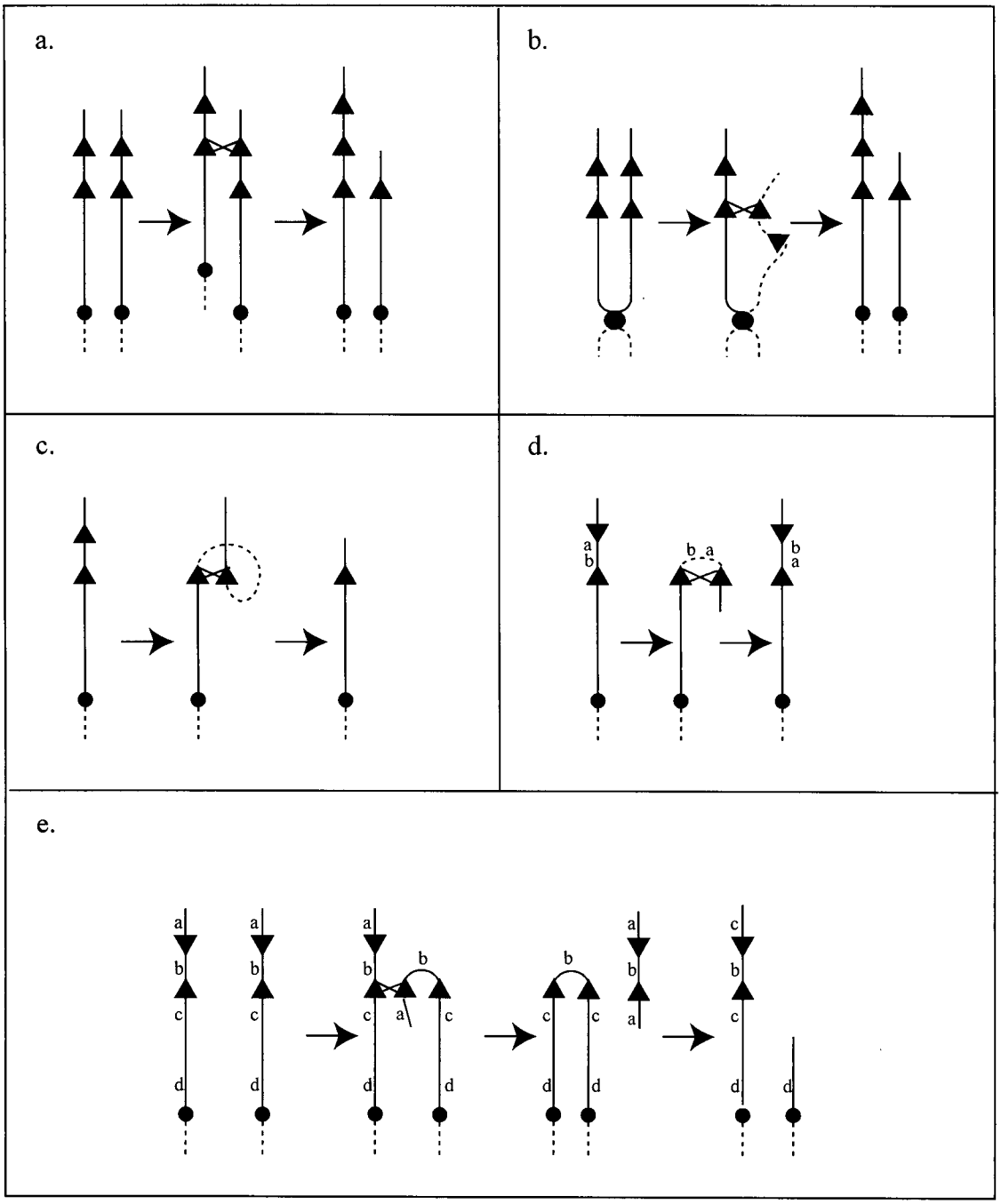


Figure 1: Proposed mechanisms for generation of recurrent chromosome rearrangements. Modified from Robinson *et al.*, 1998. Illustrations are schematic representations of chromosomal interactions. Triangles represent repeats. Dotted lines at regions involved in pairing represent the 3-dimensional aspect of pairing. All mechanisms rely on mispairing at repetitive elements, followed by recombination and resolution. Pairing may occur between direct repeats (a, b, c) or inverted repeats (d, e), and may occur between homologues (a, e), between sister chromatids (b), or within a single chromatid (c, d). Products may be duplications (a, b), deletions (a, b, c, e), inversions (d) and inversion duplications (e).

Misalignment of homologues or sister chromatids, at repetitive sequences present in direct orientation, followed by recombination at these sequences, would lead to duplication and deletion chromosomes (figure 1a, 1b). However, if the misalignment occurred within a single chromatid (figure 1c), a deletion chromosome, but not the reciprocal duplication chromosome, would be recovered. These mechanisms have been implicated in the formation of the Charcot-Marie-Tooth type 1A (CMT1A) and hereditary neuropathy with liability to pressure palsies (HNPP) chromosomes. CMT1A arises from a duplication of a 1.5 Mb region of 17p11.2-p12 flanked by homologous repetitive DNA (CMT1A-REP) whereas HNPP arises from a deletion of the same region (Lupski *et al.*, 1991, Raeymaekers *et al.*, 1991). Initially, it was suggested that these chromosomes may be the reciprocal events of an interchromosomal recombination (Chance *et al.*, 1994). However, a recent study suggests that two distinct sex-dependent mechanisms are involved in the formation of these chromosomes (Lopes *et al.*, 1998). Paternal duplications and deletions result from interchromosomal recombination at misaligned CMT1A-REP sequences but maternal duplications and deletions arise from intrachromosomal recombination, between sister-chromatids or within a single chromatid, at misaligned CMT1A-REP sequences (Lopes *et al.*, 1998). Other examples of deletion resulting from intra- or inter-chromosomal recombination at misaligned repetitive sequences include deletions of 7q11.23 implicated in William syndrome (Urbàn et al., 1996), deletions of the steroid sulfatase gene on distal chromosome Xp (Yen et al., 1990), and some deletions of the α-globin gene cluster on chromosome 16p13.1 (Nicholls et al., 1987).

Misalignment within a single chromatid, at repetitive sequences present in inverse orientation, followed by recombination at these sequences, would lead to an inversion of the region between

the repetitive sequences (figure 1d). This mechanism has been implicated in the cause of 13% of Hunter Syndrome (HS) cases. In this instance recombination has occurred between the iduronate-2-sulphatase (IDS) gene on Xq28 (Wilson *et al.*, 1993, Malmgren *et al.*, 1995) and an IDS-related region (IDS-2) present 90 kb telomeric and in the inverse orientation to the IDS gene (Bondeson *et al.*, 1995). The result is an inversion of the region causing an interruption of the IDS gene. Further, the recombination event may occur within the same highly homologous region of IDS and IDS-2 in all patients (Bondeson *et al.*, 1995), implying that these regions may be prone to recombination. This mechanism is also proposed to account for 50% of cases of severe haemophilia A. Recombination between the factor VIII gene located at Xq28 and an upstream associated gene present in inverse orientation leads to interruption of the factor VIII gene resulting in severe haemophilia A (Lakich *et al.*, 1993). It has been suggested that the proximity of factor VIII to the telomere may decrease steric constraints on intrachromosomal pairing (Lakich *et al.*, 1993). It is interesting that both of these examples of inversion resulting from intrachromosomal recombination within a single chromatid occur in the same region of the genome (Xq28). Perhaps there are features of this region of the genome that predispose to this type of interaction.

Proximal chromosome 15q is prone to a variety of recurrent rearrangements. Prader-Willi syndrome (PWS) and Angelman syndrome (AS) result from interstitial deletion of this region (Robinson *et al.*, 1991, Mascari *et al.*, 1992, Zackowski *et al.*, 1993, Saitoh *et al.*, 1994). This region is also prone to interstitial duplication (Clayton-Smith *et al.*, 1993, Mutirangura *et al.*, 1993, Abeliovich *et al.*, 1995) and to the formation of isodicentric inversion duplication supernumerary chromosomes (discussed below) (Cheng *et al.*, 1994, Leana-Cox *et al.*, 1994, Robinson *et al.*, 1997). Although the breakpoints of these rearrangements differ, both proximal

and distal breakpoints cluster near a series of repeated sequences (Robinson *et al.*, 1997, Buiting *et al.*, 1992, Amos-Landgraf *et al.*, 1994). It has been suggested that these rearrangements may arise due to inter- or intra-chromosomal misalignment and recombination at these sequences (Robinson WP, personal communication). The difference in breakpoints may reflect differences in interactions between these sequences dependent on location and orientation (Robinson et al., 1998).

1.4 Inversion Duplications

Misalignment of homologues at repetitive sequences present in inverse orientation, followed by recombination at these sequences, would lead to a dicentric chromosome and an acentric fragment (figure 1e). Resolution of this event through breakage of the dicentric chromosome could lead to an inverted duplication deficiency chromosome and the reciprocal deletion chromosome. The existence of repetitive sequences at sites prone to this sort of rearrangement has not been shown.

Inversion duplication chromosomes have been described as mirror duplication chromosomes because standard G- or R-banding techniques give a mirror image appearance to the regions involved in the duplications. This is due to a duplication of a portion of the chromosome present in the opposite orientation to the original segment so that the bands are reflected out from a center of symmetry. Inversion duplications may be interstitial, or terminal with a deletion of distal sequences, and invariably result in an unbalanced karyotype. An example of an inversion duplication with distal deletion is shown in figure 2. The inversion duplication 8p (inv dup(8p)) chromosome in this example has a duplication from band 8p12 to 8p23.1 and a deletion from band 8p23.1 to the telomere. The duplicated material is present at the end of the aberrant

chromosome in opposite orientation to the original segment, with a center of symmetry at 8p23.1.

In some examples, the inversion duplication chromosome is present as a supernumerary chromosome, also known as supernumerary pseudodicentric chromosomes. Most of these chromosomes are derived from chromosomes 15 (Cheng *et al.*, 1994, Robinson *et al.*, 1993, Wandstrat *et al.*, 1998) or 22 (Mears et al., 1994) and present as bisatellited marker chromosomes comprised of the p arm material plus a small segment of q arm material. These chromosomes may be asymmetrical, are usually dicentric, and are usually found in unrelated patients. The chromosome 15 supernumerary inversion duplication (inv dup(15p)) chromosomes containing the PWS/AS region can be classified into at least two categories based on similar breakpoint locations. These supernumerary inv dup(15p) chromosomes arise by an interchromosomal maternal event (Wandstrat *et al.*, 1998), that must be followed by non-disjunction (Schrek *et al.*, 1977). It has been suggested that variable interactions at several low-copy repeat sequences present in this region could lead to the formation of different classes of inv dup(15p) chromosomes (Wandstrat *et al.*, 1998). Alternatively, U-type recombination between homologues (Van Dyke, 1988) may be involved.

Inversion duplication chromosomes are rare in live-born individuals. Van Dyke (1988) observed that of the 20 inversion duplication cases, excluding patients with supernumerary chromosomes, found in the literature at that time, there was an over-representation of those involving 4q, 9p, and 8p. The number of reported cases of inversion duplication 8p has increased to over 60 (table 1), however, this same increase in number has not been reported for 4q, 9p, or any other chromosome location. The center of symmetry of the inversion duplication 8p chromosomes is similar in all patients. In the majority of cases the center of symmetry is

within 8p22 or 8p23 (figure 2), and the duplication extends to at least band 8p21.2. This identity of center of symmetry in unrelated cases is also seen in a few rarer cases of inversion duplication, including Xp inversion duplications (Tuck-Muller *et al.*, 1993, Telvi *et al.*, 1996), 7q inversion duplications (Stetten *et al.*, 1997) and chromosome 14 inversion duplications (North *et al.*, 1995). For these rarer examples the number of cases is extremely small, usually only two, and it is therefore difficult to determine whether this is an ascertainment bias based on viability of the rearrangement or represents some underlying feature of the genome at these locations. However, at least for chromosome 8p, there appears to be a propensity to form this type of rearrangement. This led to investigation of whether other features of the inv dup(8p) chromosomes were shared, and whether this may reflect a common mechanism of origin.

1.5 Inversion Duplication 8p (inv dup(8p))

The inversion duplication 8p chromosome has been described above. The majority of these chromosomes have a duplication involving at least bands 8p12 to 8p23, and a center of symmetry at 8p23.1 or 8p22 (figure 2). In 1987, Dill *et al.* reported the deletion of bands distal to the center of symmetry, for the first time describing these chromosomes more correctly as inversion duplication deficiency chromosomes. Dicentric inv dup(8p) chromosomes have also been reported (Floridia *et al.*, 1996). These chromosomes have a single q arm, an inverted duplication of the region from the centromere to band 8p23.1, a deletion of distal bands, and centromere associated sequences, such as alphoid DNA, present near the telomere of the p arm. These dicentric chromosomes represent the largest duplication ascertained for inv dup(8p), the smallest being from band 8p21.2 to band 8p23.1. All other inv dup(8p) chromosomes vary in the amount of duplication, within this range.

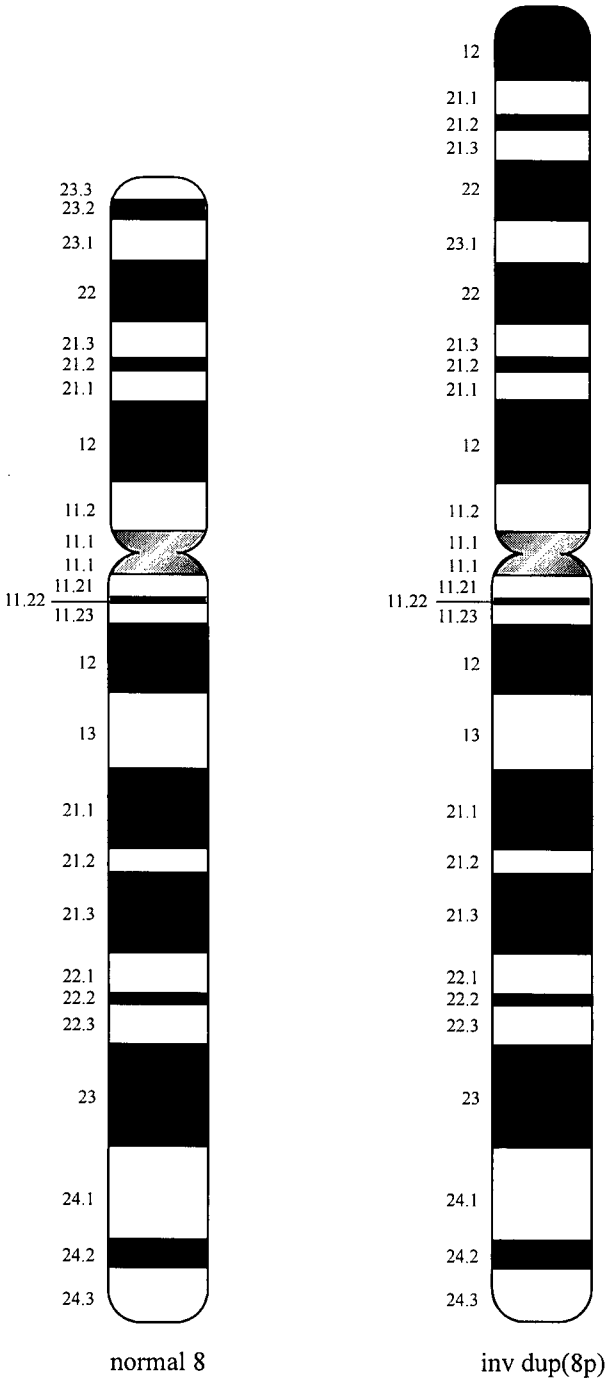


Figure 2: Ideogram of normal chromosome 8 and inversion duplication (8p) chromosome.

Table 1: A comprehensive list of reported cases of inversion duplication (8p) patients.

The reported region of duplication is listed, with the center of symmetry being the second band location. Report of a distal deletion is indicated by 'yes', report of an undetected distal deletion is indicated by 'no'. The origin of the chromosome is indicated (origin) as well as whether the parental DNA from which the chromosome was derived is known (DNA origin). *d.n.* = *de novo*, rec. = recombinant chromosome, n.k. = not known, mat. = maternal origin

Reference (et al.) & Year	**Case**	**Duplicated Region**	**Origin**	**Distal Deletion?**	**DNA Origin**
Weleber 76	1	8p11.2→p23.1	d.n.	n.k.	n.k.
Taylor 77	1	8p12→8p23	d.n.	n.k.	n.k.
Taylor 77	2	8p21→8p23	d.n.	n.k.	n.k.
Rethoré 77	1	8p11→8p22	d.n.	n.k.	n.k.
Rethoré 77	2	8p11→8p22	n.k.	n.k.	n.k.
Hongell 78	1	8p12→8p23?	d.n..	n.k.	n.k.
Mattei 80	1	8p12→8p23.2	n.k.	n.k.	n.k.
Mattei 80	2	8p11→8p23.2	n.k.	n.k.	n.k.
Poloni 81	1	8p11→8p23.2	n.k.	n.k.	n.k.
Poloni 81	2	8p11→8p23.1	n.k.	n.k.	n.k.
Poloni 81	3	8p11→8p23.1	n.k.	n.k.	n.k.
Jensen 82	1	8p21.2→8p23.1	d.n.	n.k.	n.k.
Jensen 82	2	8p21.1→8p23.3	d.n.	n.k.	n.k.
Fryns 85 /Kleczkowska 87	1/3	8p21.1→8p22	d.n.	n.k.	n.k.
Kleczkowska 87	1	8p21.1→8p22	d.n.	n.k.	n.k.
Kleczkowska 87	2	8p21.1→8p22	d.n.	n.k.	n.k.
Dill 87/Henderson 92	G.S.	8p12→8p23.1	d.n.	yes	n.k.
Nevin 90	1	8p12→8p23.1	d.n.	n.k.	n.k.
Gorinati 91/Minelli 93/ Floridia 96	1/3/14	8p21.1→8p23.1	d.n.	yes	n.k.
Feldman 93	1	8p21→8p23	d.n.	n.k.	n.k.
Feldman 93	2	8p21→8p23	rec.	n.k.	mat.
Feldman 93	3	8p11.23→8p23.1	d.n.	n.k.	n.k.
Feldman 93	4	8p11.2→8p23	d.n.	n.k.	n.k.
Feldman 93	5	8p21.1→8p23.1	d.n.	n.k.	n.k.
Feldman 93	6	8p11.2→8p23.1	d.n.	n.k.	n.k.
Feldman 93	7	8p11.2→8p23.1	d.n.	n.k.	n.k.
Feldman 93	8	8p12→8p23	d.n.	n.k.	n.k.
Feldman 93	9	8p12→8p23	d.n.	n.k.	n.k.
Feldman 93	10	8p12→8p23.1	d.n.	n.k.	n.k.
Minelli 93/Floridia 96	1/10	8p12→8p22/ 8p11.2→p23.1	d.n.	yes	mat.

Table 1 continued

Reference (et al.) & Year	**Case**	**Duplicated Region**	**Origin**	**Distal Deletion?**	**DNA Origin**
Minelli 93/Floridia 96	2/2	8p12→8p22/ 8cen→8p23.1	d.n.	yes	mat.
Minelli 93/Floridia 96	4/7	8p12→8p22/ 8p11.2→8p23.1	d.n.	yes	mat.
Minelli 93/Floridia 96	5/8	8p12→8p22/ 8p11.2→8p23.1	d.n.	yes	mat.
Minelli 93/Floridia 96	6/9	8p12→8p22/ 8p11.2→8p23.1	d.n.	yes	mat.
Minelli 93/Floridia 96	7/16	8p12→8p22 /8p21→8p23.1	d.n.	yes	mat.
Minelli 93/Floridia 96	8/5	8p12→8p22/ 8cen→8p23.1	d.n.	yes	mat.
Minelli 93/Floridia 96	9/1	8p12→8p22/ 8cen→8p23.1	d.n.	yes	mat.
Barber 94	1	8p11.23→8p23.1	d.n.	yes	mat.
Mitchell 94	1	8p12→8p23.1	d.n.	yes	n.k.
Engelen 94/ de Die-Smulders 95	1/1	8p12→8p23.1	d.n.	yes	n.k.
Engelen 94/ de Die-Smulders 95	2/2	8p21.1→8p22	d.n.	yes	n.k.
Engelen 94/ de Die-Smulders 95	3/6	8p12→8p23.1	n.k.	yes	n.k.
Engelen 94/ de Die-Smulders 95	4/7	8p12→8p23.1	n.k.	yes	n.k.
Redha 94	1	8p21→8pter	n.k.	n.k.	n.k.
Hoo 95	3	8p21.2→8p23.2	d.n.	n.k.	n.k.
de Die-Smulders 95	3	8p21.1→8p22	d.n.	n.k.	n.k.
de Die-Smulders 95	4	8p11.2→8p23.1	d.n.	n.k.	n.k.
de Die-Smulders 95	5	8p12→8p23.1	d.n.	n.k.	n.k.
Guo 95	1	8p12→8p23.1	d.n.	yes	n.k.
Guo 95	2	8p12→8p23	d.n.	yes	n.k.
Guo 95	3	8p12→8p23	n.k.	yes	n.k.
Guo 95	4	8p11.2→8p23.2	d.n.	n.k.	n.k.
Guo 95	5[a]	8p21.3→8p23.3	n.k.	n.k.	n.k.
Guo 95	6	8p11.2→8p23.1	n.k.	n.k.	n.k.
Guo 95	7	8p11.2→8p23.1	d.n.	n.k.	n.k.
Floridia 96	3	8cen→8p23.1	d.n.	yes	mat.
Floridia 96	4	8cen→8p23.1	d.n.	yes	mat.
Floridia 96	6	8cen→8p23.1	d.n.	yes	mat.
Floridia 96	11	8p11.2→8p23.1	d.n.	yes	mat.
Floridia 96	12	8p11.2→8p23.1	d.n.	yes	mat.

Table 1 continued

Reference (et al.) & Year	**Case**	**Duplicated Region**	**Origin**	**Distal Deletion?**	**DNA Origin**
Floridia 96	13	8p11.2→8p23.1	d.n.	yes	mat.
Floridia 96	15	8p21→8p23.1	d.n.	yes	mat.
Nelson, this thesis	TP	8p11.2→p23.1	d.n.	yes	mat.
Nelson, this thesis	DC	8p12→p23.?2	d.n.	no	n.k.
Nelson, this thesis	SW	8p21.3→p23.?2	d.n.	no	n.k.
Nelson, this thesis	MD[b]	8p+	d.n.	yes	mat.

[a] Patient was diagnosed as 47,XX, inv dup(8)(p23.3p21.3)?, +r(?). The abnormal banding pattern on the derivative 8p is similar to other cases (Guo *et al.*, 1995)

[b] Patient carries an inv dup(8p) chromosome with satellites, see section 3.1.2

The phenotype of patients who have an inv dup(8p) chromosome is variable. This has been attributed to three potential causes: variation in the extent of duplicated material, variation in genetic background, and variation in the extent of deleted material. However, deletion of distal bands may not have a significant role in the phenotype of these patients as deletions of terminal (8p) have only a mild phenotypic effect (Fryns *et al.*, 1989). In an attempt to define common features of the phenotype for inv dup(8p) patients, Kleczkowska *et al.* (1987) studied patients with inverted duplications of bands 8p21.1→8p22. Some common features seen in these patients include severe mental retardation, skeletal problems, difficulties sucking and feeding as infants, and a characteristic facies with features including: quadrangular head, prominent forehead, malformed and posteriorly rotated ears, thin lips with everted lower lip, broad nasal bridge with prominent nose and anteverted nostrils.

1.5.1 Inversion Duplication 8p is Sporadic

To date, none of the inv dup(8p) chromosomes has been inherited from a parent carrying the same aberrant chromosome. This is not surprising as it is rare for individuals with severe mental retardation to reproduce. Feldman *et al.* (1993) reported an inv dup(8p) patient whose mother carried a balanced paracentric inversion. The child's inv dup chromosome is believed to have arisen as a result of recombination in a meiotic inversion loop. All other confirmed cases of inv dup(8p), where origin has been investigated, have been *de novo* (table 1). In 1994, Dhooge *et al.* reported a case that, based on cytogenetic analysis, may be an inversion duplication or a direct duplication, segregating in a family. The mother and the two children carrying the aberrant chromosome have mild mental retardation. Based on molecular and fluorescence *in situ*

hybridization (FISH) evidence presented in this thesis, this rearrangement appears to be a direct duplication.

1.5.2 Center of Symmetry of Inv Dup(8p)

In order to define inversion duplication (8p) as a recurrent rearrangement, common features of the inv dup(8p) chromosome must be seen in unrelated patients. Although the amount of duplication varies from patient to patient, the cytogenetic center of symmetry is reported to be in band 8p22 or band 8p23.1 in the majority of patients (see table 1), with rare cases in band 8p23.2 or 8p23.3 (Mattei *et al.*, 1980, Poloni *et al.*, 1981, Jensen *et al.*, 1982, Redha *et al.*, 1994, Hoo *et al.*, 1995, Guo *et al.*, 1995). However, a number of reports of inv dup(8p) chromosomes with a center of symmetry at 8p22 were made prior to either the routine use of high resolution chromosome banding techniques (Rethoré *et al.*, 1977), or before it was recognized that there is a deletion of DNA distal to the center of symmetry (Kleczkowska *et al.*, 1987, Fryns *et al.*, 1985). In some cases, re-examination of inv dup(8p) chromosomes has led to redefinition of the center of symmetry to band 8p23.1 (Gorinati *et al.*, 1991 and Minelli *et al.*, 1993 re-examined in Floridia *et al.*, 1996). Therefore, it can be argued that in some, if not all, instances where the center of symmetry is reported in band 8p22, the correct interpretation should be 8p23.1. Thus, with rare exceptions, the majority of inv dup(8p) chromosomes have a common center of symmetry in band 8p23.1.

1.5.3 Molecular Features of Inversion Duplication 8p

As well as the common phenotypic features and similarity in the cytogenetic location of the center of symmetry in unrelated inv dup(8p) patients, there are a number of common molecular

features of these chromosomes. A deletion of distal material on the inv dup(8p) chromosome was first shown at the molecular level by Dill *et al.* (1987) by densitometric Southern blot analysis, using a probe for the D8S7 locus mapping to 8p23 (Wood *et al.*, 1986). The patient DNA was monosomic for the marker, which indicated transmission of only one parental allele at the D8S7 locus, and lead to the conclusion that the aberrant inv dup(8p) chromosome must be a duplication-deficiency chromosome. This duplication-deficiency was later confirmed by FISH analysis (Henderson *et al.*, 1992). Approximately half of the reported cases have subsequently been tested for deletion of distal sequences, using either molecular techniques or FISH analysis (Barber *et al.*, 1994, Mitchell *et al.*, 1994, de Die-Smulders *et al.*, 1995, Guo *et al.*, 1995, Floridia *et al.*, 1996). In all informative cases, where the center of symmetry is in 8p23.1 or 8p22, terminal deletions have been found (see table 1). The most proximal marker deleted from inv dup(8p) chromosomes is D8S349, located within band 8p23.1 (see figure 3) (Floridia *et al.*, 1996). The inv dup(8p) chromosome is derived from both maternal chromosomes in 17/17 cases investigated (Minelli *et al.*, 1993, Feldman *et al.*, 1993, Floridia *et al.*, 1996). Therefore, these chromosomes must arise by a maternal interchromosomal recombination event.

Using a combination of genotyping and/or FISH analysis, Floridia *et al.* (1996) defined a region located at the center of symmetry called the single copy region. In all informative cases, this region is found in band 8p23.1, delimited at the distal end by D8S349 (deleted from the aberrant chromosome) and at the proximal end by D8S552 (duplicated on the aberrant chromosome) (see figure 3 for locations). The sequence tag site (STS) markers located within this region are present in single copy on the aberrant chromosome, and are flanked by regions of duplication. Therefore, molecular evidence suggests that these chromosomes are asymmetrical.

In summary, the majority of informative cases of individuals with an inv dup(8p), the aberrant chromosome has a center of symmetry located at 8p23.1, a deletion of genetic material distal to the center of symmetry, a region of single copy located at the center of symmetry, and is derived from a maternal interchromosomal event. The finding of shared features of the aberrant chromosome in unrelated patients suggests that these features may reflect a common mechanism of origin.

1.5.4 Mechanism of Formation of inv dup(8p)

Two mechanisms proposed to account for the common features of the inv dup(8p) chromosomes are presented here:

The first mechanism (Weleber *et al.*, 1976) is the end-to-end fusion of two broken chromosomes 8 forming a dicentric chromosome which would break at anaphase to create the inversion duplication chromosome and the reciprocal deletion chromosome. Based on the observation in inv dup(8p) chromosomes of a distal deletion, common cytogenetic 'center of symmetry', and region of single copy, a refinement of this mechanism has been made (Dill *et al.*, 1987, Floridia et al., 1996). A mechanism involving inverted repetitive sequences, analogous to that proposed for other recurrent structural rearrangements, could apply to inv dup(8p). Misalignment of homologues, at these inverted sequences, followed by recombination, would result in a dicentric chromosome (figure 4). As the chromosomes proceed through normal anaphase, the centromeres of the dicentric chromosome may be pulled to opposite poles, causing the dicentric to break at a random point between, or at, the centromeres. It is widely accepted that telomere repair acts to heal broken ends (Morin 1991, Flint *et al.,* 1994) and therefore, could repair the broken ends and stabilize the aberrant products. Of the resulting daughter cells, one

would contain an inversion duplication chromosome, the other, a deletion chromosome. The amount of duplication on the inversion duplication chromosome would vary, depending on the location of the break. Furthermore, if the paired repeat sequences are located relatively far apart, then a region of single copy, flanked by the duplicated regions, would be detected on the aberrant chromosome (Floridia *et al.*, 1996). This mechanism is particularly attractive as it requires only the naturally occurring secondary consequences of an initial aberrant pairing event, and is directly testable at the molecular level.

The second involves a U-type exchange within a paracentric inversion loop during meiosis (Mitchell *et al.,* 1994) (figure 5). In a paracentric inversion with a distal break point, a U-type exchange within the meiotic inversion loop would leave the telomeres intact while deleting subtelomeric sequences. This mechanism requires that a parent carry the inversion and could lead to familial cases of inv dup(8p). When both parents are karyotypically normal, however, only a premeiotic event creating a new inversion would allow this type of mechanism to occur. Since this would require two rare events, inversion and U-type exchange, and since this exchange would not lead to a region of single copy at the center of symmetry of the aberrant chromosome, it is unlikely that this mechanism would account for the majority of cases reported. All parents of inv dup(8p) patients, with the exception of the mother of the Feldman *et al.* (1993) case, are karyotypically normal.

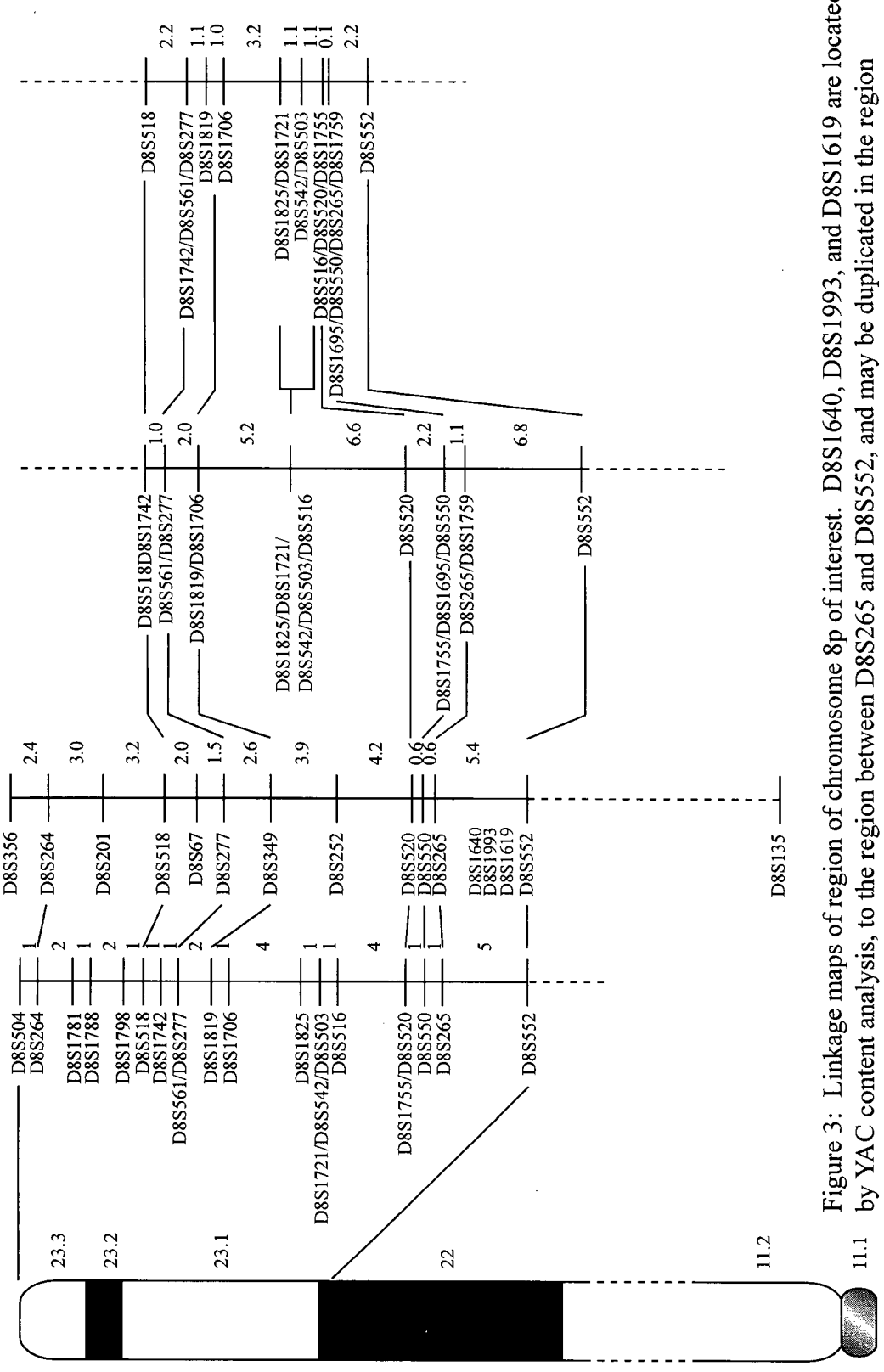


Figure 3: Linkage maps of region of chromosome 8p of interest. D8S1640, D8S1993, and D8S1619 are located, by YAC content analysis, to the region between D8S265 and D8S552, and may be duplicated in the region between D8S349 and D8S252.

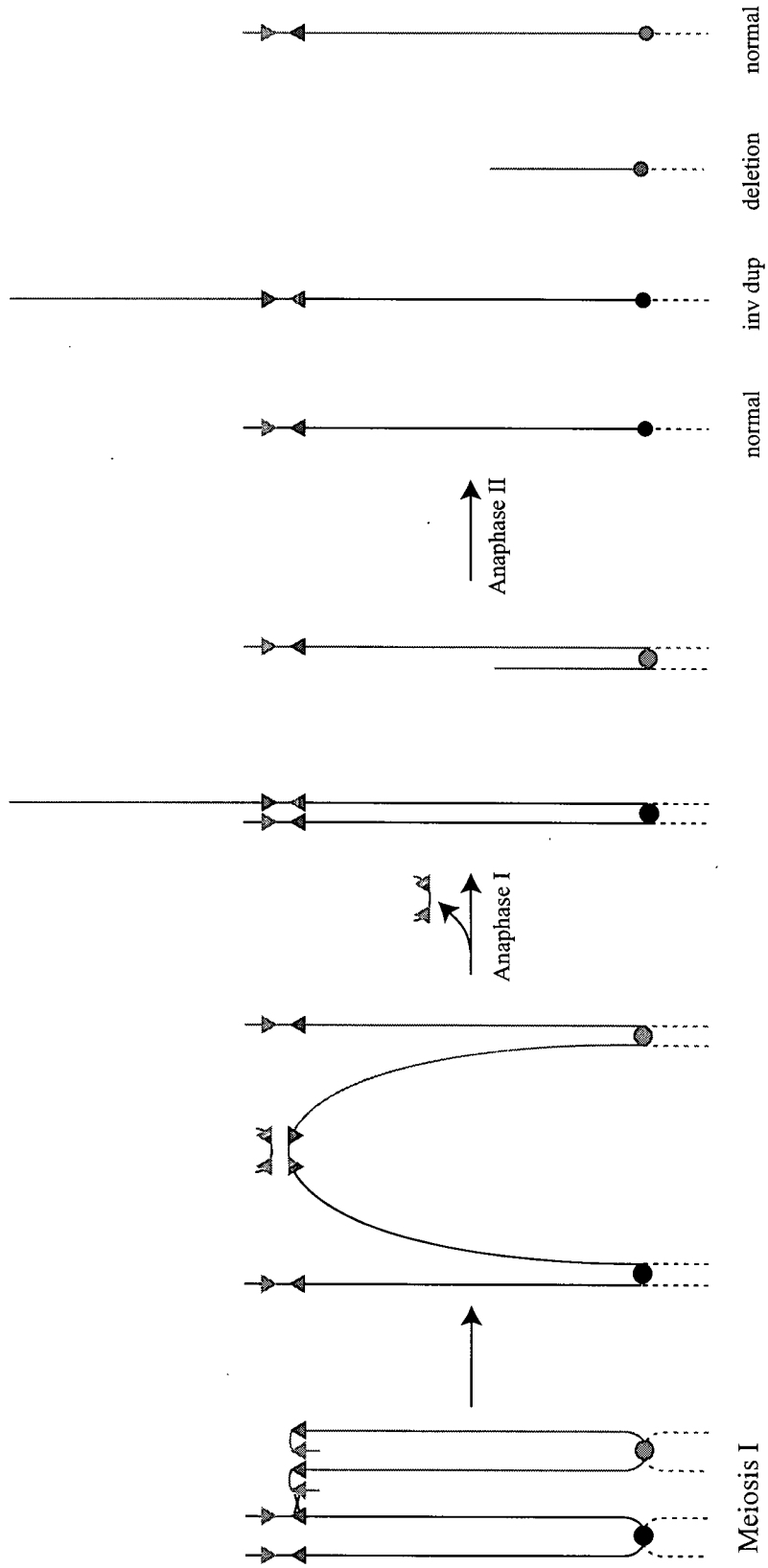


Figure 4: A proposed mechanism for the generation of inversion duplication(8p). Illustration is a schematic representation of chromosome interactions. Aberrant pairing at inverted repetitive sequences, followed by recombination at the repetitive sequences, generates a dicentric and acentric fragment. Breakage of the dicentric chromosome, followed by telomere repair results in an inversion duplication chromosome, the reciprocal deletion chromosome, and two normal chromosomes.

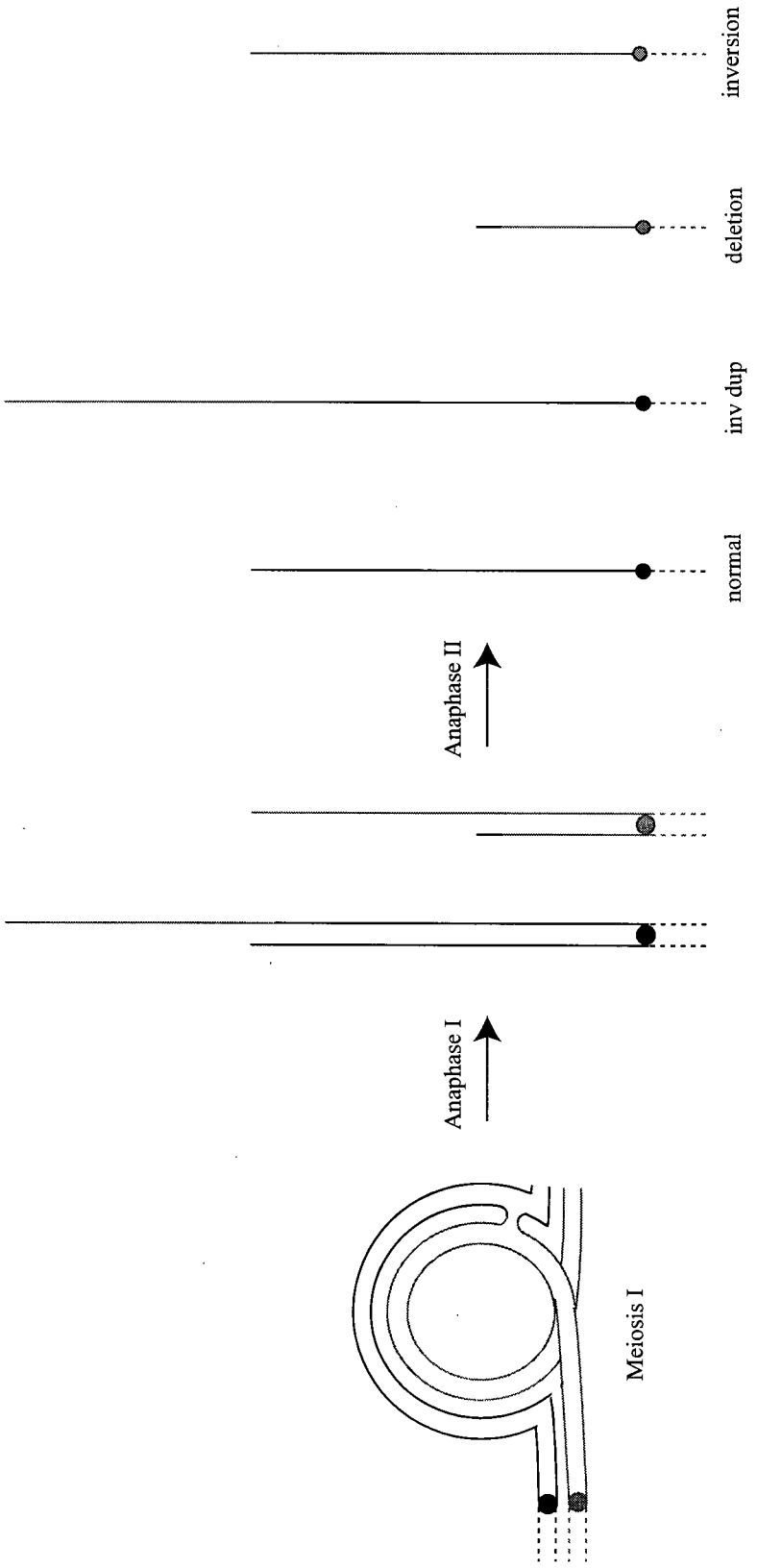


Figure 5: A proposed mechanism for the generation of inversion duplication(8p). Illustration is a schematic representation of chromosome interactions. A U-type exchange occurs within a paracentric inversion loop. Resolution generates an inversion duplication chromosome as well as the reciprocal deletion chromosome and the original normal and inversion chromosomes.

1.6 Project Objectives

The objective of this thesis was to investigate inversion duplication (8p) chromosomes at the molecular level to determine whether a mechanism mediated by inverted repeats could apply. To accomplish this objective, molecular genetic analysis was carried out on inversion duplication (8p) patients, with the intent of isolating proximal and distal 'breakpoint' sequences and examining these sequences for similarity. The breakpoint is defined as the point at which the event has occurred, in this case, by the transition from one maternal chromosome 8 homologue to the second maternal chromosome 8 homologue. Should a mechanism mediated by repetitive sequences apply, this transition point would be located either distal or proximal to the single copy region, depending on where the proposed recombination event occurred. Further, should repetitive sequences be found at these transition points, these sequences should be inverted with respect to one another.

Genotyping of inv dup(8p) patients for genetically and physically mapped STS markers allowed the distal breakpoint to be localized to a region flanked by STS markers D8S349 and D8S503 (figure 3). The proximal breakpoint location is defined by STS marker D8S265 and D8S552 (Floridia *et al.*, 1996). As a first step towards isolating the breakpoint sequences, fine-scale physical and genetic mapping of these regions was required. Mapping efforts were focused on the distal region.

A novel repetitive element identified by Kogi *et al.*, 1997 and Gondo *et al.*, 1996 on chromosome 4 was isolated and mapped to the regions of 8p containing the proximal and distal breakpoints. Therefore, this element is a candidate for the proposed repetitive element involved in the formation of inv dup(8p).

Chapter 2: Materials and Methods

2.1 Polymerase Chain Reaction

The annealing temperature of each primer of the sequence tag site (STS) primer pair was calculated by the equation (4(C+G) + 2(A+T) - 4). The annealing step of PCR was carried out at the lowest calculated annealing temperature of the pair, allowing specific hybridization of both primers. Sequence and annealing temperature of primer pairs used in this thesis are listed in table 2.

100 ng of sample DNA was amplified in a 25 µl reaction volume containing 2.5 mM $MgCl_2$, 0.2mM of each nucleotide (dATP, dTTP, dCTP, dGTP), 50 mM Tris-Cl pH 8.3, 0.05% Nonidet-40, 0.05% Tween-20, 0.4 mM each primer pair and 0.5U *Taq* Polymerase. A positive control of 100 ng total human DNA and a negative control of Milli Q filtered distilled water were included with each set of reactions. Samples were overlaid with 25 µl of paraffin oil to prevent evaporation. An automated thermal cycler was used to amplify the DNA. Forty cycles of denaturation (1 min at 95° C), annealing (30 seconds at the appropriate annealing temperature) and elongation (1 min at 72° C) were carried out, followed by a 10 minute elongation step at 72° C. 5 µl of stop buffer was added prior to gel electrophoresis.

Stop Buffer
0.25% xylene cyanol
0.25% bromophenol blue
40% sucrose (w/v)
60 mM EDTA

Table 2: STS marker information.

T_A= annealing temperature, Max. Hetero. = Maximum Heterozygosity, n/a = not applicable

Locus Symbol	primer sequences (5′-3′)	product size (bp)	T_A (°C)	polymorphic	Max. Hetero.	References
D8S201	ATCAGACCAATAACCCCAGG	170-285	55	yes	0.92	Tomfohrde *et al.*, 1992
	TGGCTAACACGGTGAAACCA					
D8S349	AGTAGGCCTGGAAGGACTTT	115-133	56	yes	0.85	Weber, unpublished (1993)
	TGCTGAGAATGAGACCCATC					
D8S503	GACCATGATTAAGCAAAACAAA	212-226	50	yes	0.74	Gyapay *et al.*, 1994
	TCGCTCAGAAACAAACCAA					Weissenbach, unpublished (1993)
D8S265	ACCTCTTTCCAGATAAGCCC	208-231	54	yes	0.78	Gyapay *et al.*, 1994
	CCAATGGTTTCGGTTACTGT					Weissenbach, unpublished (1992)
D8S552	AGGATTGTAATTTCCTTGC	168-182	50	yes	0.79	Gyapay *et al.*, 1994
	GGGACTTTTTGAAGGTTTG					Weissenbach, unpublished (1993)
D8S136	GCCCAAAGAGGAGAATAAA	71-89	58	yes	0.881	Mitchell & Wood, unpublished (1992)
	CGAAGCCACACCTTTGTC					
D8S135	ACCACCTAAAACCCTGTTCAC	150-184	58	yes		Wood, et al. (1995)
	CATGGTTGCGGGGAGTCAG					
STS4-310	CTCATTAGGGAAGAAAATACGCTG	251	56	no		Goold *et al.*, 1993
	CAATGGCAGGGAAGAAATCTTTA					
SHGC4-672	TCGCCACCTTGAATAGTGGAAAC	180	62	no		Goold *et al.*, unpublished (1994)
	TGAAACAGGCAAGTCGGTCTGTC					

Table 2 continued

Locus Symbol	primer sequences (5′-3′)	product size (bp)	T_A (°C)	polymorphic	Max. Hetero.	References
SHGC-1030	AGACCTTCCAGCGATAACCTTTC	166	62	no		Myers, unpublished (1994)
	TCAAACTTCAGTGCATGTTTCTACC					
SHGC4-1135	CACGGCATTTCAGACACTTTTGAC	142	62	no		Goold *et al.*, unpublished (1994)
	TGGAGACAAACACCACAATGATAG					
SHGC4-1596	AGTGTTCATGGACTCCTGATGTG	161	62	no		Goold *et al.*, unpublished (1994)
	AGTTCAACGTCAGAAAGTCGAAG					
SHGC4-1739	CCTGTGCACTCCTTCCTGAATTG	107	56	no		Goold *et al.*, unpublished (1994)
	AGCCCTTCTGTAAGTCACCCCC					
SHGC-1436	GTAAGGAACAGGGCGGAATTGAG	160	62	no		Goold *et al.*, unpublished (1994)
	CCATGAAAAGACAAGGGAAAGAAAC					
SHGC4-1656	CAGAAGCATGAGCAATGCAGATG	173	62	no		Myers, unpublished (1995)
	GTGTGTGAAAACTACAGTGTGATG					
rsAVA13	CTCATTAGGGAAGAAAATACGCTG	152	56	no		Vocero-Akbani *et al.*, 1996
	TCTCTGGCCAAATCAGCC					
D8S1819	TCACTGAGGGACTTGGC	207-223	50	yes	0.82	Dib *et al.*, 1996
	CGTGCTGAGAATGAGACC					Weissenbach, unpublished (1996)
D8S277	TGAGAGGTCTGAGTGACATCCG	148-180	64	yes	0.73	Gyapay *et al.*, 1994,
	CCAGGTGAGTTTATCAATTCCTGAG					Weissenbach *et al.*, 1992
D8S439	GGCTCCCTGTTCTTTATCAGGTTG	115-133	60	yes	0.5	Gerken, 1993
	ACTTTTCTCTGGCTATTATGGACTC					

Table 2 continued

Locus Symbol	primer sequences (5′-3′)	product size (bp)	T_A (°C)	polymorphic	Max. Hetero.	References
D8S1935	CTTCACAAAGAGAAAAATACAAGCC	328	55	no		Hudson *et al.*, 1995
	CCAGCCTCTTTTTCTTTCCC					
D8S252	ATTTCTCTCCACACTGACATAC	190	58	yes	0.27	Wood, unpublished (1992)
	CGTGGTTCTTACAGGAGGGTTA					
D8S1640	TGCAGTCTGCGGGAGTTC	175	55	no		Hudson, unpublished (1995)
	AGCAGGGTGACTGTAAAGAAGG					
D8S1993	GCTGCTTCACCACTCAGGTC	174	55	no		Hudson, unpublished (1995)
	AGCCAGAGAATAAATGAATAAACAA					
D8S1619	GTGGTGCAGTTCATCCTCTG	226	55	no		Hudson, unpublished (1995)
	CCTTGCAAAGTATTTGGTACTAAGA					
D8S574	CCCTGTGGAGAAAGAGAGAAA	162	58	yes	0.72	Wood, unpublished (1993)
	TCCTAACTAAGTCGTGTTGGC					
D8S1825	GAGATGGGGTTTCTCTATGTTGC	139	55	yes	0.72	Dib *et al.*, 1996
	TGGGATTTCATTTTTAACCTGTG					
D8S351	AGCCAGAAATTGAGGAAGTG	109-127	55	yes	0.84	Weber, unpublished (1993)
	CTGCAAGCTCTTTCAGTTGA					
D8S516	GAGAATGCTTGACCCCAAAAAATC	158-168	62	yes	0.58	Gyapay *et al.*, 1994
	CCTAAGAGAGTGCTATGTGCTCCC					

Table 2 continued

Locus Symbol	primer sequences (5′-3′)	product size (bp)	T_A (°C)	polymorphic	Max. Hetero.	References
D8S542	AATCACCTANACTACTGCCA	229-249	55	yes	0.6	Gyapay *et al.*, 1994
	ATCTGATGGGGAGTTATGTATTC					
primer 224	CGAATCGTAACCGTTCGTACGAGAATCGC T	n/a		n/a		Riley et al., 1990
primer T7	TAATACGACTCACTATAGGGAGA	n/a	42	n/a		Biosynthesis cat#51-7000
primer SP6	ATTTAGGTGACACTATAG	n/a	42	n/a		Custom
primer ALE-1	GCCTCCCAAAGTGCTGGGATTACAG	n/a	64	n/a		Cole et al., 1991
primer ALE-3	CCACTGCACTCCAGCCTGGG	n/a	64	n/a		Cole et al., 1991

2.2 Polymorphic STS Genotyping

2.2.1 Labeling of STS Primer

The annealing temperatures of the primer pairs were calculated, and the primer with the lower annealing temperature of the pair was chosen for γ-end labeling. This decreases visualization of non-specific PCR products. Labeling was carried out in 10 µl reaction containing 50 ρmoles of DNA primer, 1X Kinase Forward Reaction Buffer (Gibco/BRL), 1U T4 Kinase (Gibco/BRL) and 25 µCi γ-dATP. Labeling was carried out at 37° C for 60 minutes followed by 5 minutes at 95° C to heat inactivate the enzyme. The primer was stored at -20° C.

2.2.2 Polymerase Chain Reaction

PCR was carried out on 100 ng of sample DNA, prepared from whole blood, using standard reactions (see section 2.1) with the exception that only 30 cycles were completed and that 0.1-0.2 µl end-labeled primer was added to each reaction tube. Upon completion of PCR, samples were frozen at -20° C. Just prior to polyacrylamide gel electrophoresis (PAGE), PAGE stop buffer was thawed and 9 µl was added to each frozen sample. Samples were boiled for 5 minutes and placed on ice. The genotype of individual 134702 (Dausset *et al.*, 1990) was included in each set of reactions as a positive control.

PAGE Stop Buffer
10 mM NaOH
95% formamide
0.05% bromophenol blue
0.05% xylene cyanol

2.3 Polyacrylamide Gel Electrophoresis (PAGE)

2.3.1 Preparation of Gel Apparatus and Casting Tray

The reservoir plate, front plate and two 0.4mm spacers were cleaned with 95% ethanol and low-lint Kimwipes. The reservoir plate was silanized and placed silene side up on a level surface. The spacers were placed along either side of the reservoir plate and covered with the front plate. The plates were held together using sealers, allowing approximately 1 cm of glass to protrude from the bottom of the sealers. The PAGE casting tray was lined with a casting tray sponge and a piece of Whatmann paper cut to the dimensions of the sponge.

2.3.2 Preparation of the Polyacrylamide Gel

5.2 g of urea was dissolved in 27 ml dH_2O, 7.2 ml 5X TBE and 6 ml Long Ranger (Mandel Scientific Corp.). To 13 ml of this solution was added 56 µl TEMED (Promega) and 200 µl 25% (w/v) ammonium persulfate (APS), swirled to mix, and poured quickly into PAGE casting tray. A seal was formed by placing the bottom of the PAGE gel apparatus onto the seal solution and allowing it to move between the glass plates by capillary action. To the remaining solution was added 26 µl TEMED and 106 µl of 25% APS, swirled to mix, and drawn up into a 25 ml plastic pipet using an automated pipeter. The gel solution was carefully poured from the pipet between the two glass plates until filled. The comb was placed flat side down between the plates and the remaining gel solution was poured over the comb to seal the area from air. The gel was allowed to polymerize for 1 hour, the casting tray was removed, and the gel placed in the PAGE running apparatus containing 0.6X TBE. The comb was removed, the reservoir filled with 0.6X TBE, the

electrodes placed, and the gel allowed to run at 3000 V, 150 mA, 55 W for 20-30 minutes to pre-warm the apparatus.

2.3.3 PAGE Electrophoresis and Autoradiography

The space made for the comb was cleaned with 0.6X TBE and all air bubbles were removed. The comb was placed and the samples were loaded. The gel was subjected to electrophoresis for the appropriate length of time (for allele sizes of approximately 130 bp this was 5 minutes after the first marker dye had run off the gel). The gel was removed from the gel apparatus onto Whatmann paper, dried in a gel dryer, placed on film for the appropriate length of time to obtain clear signal, and developed.

2.4 Cosmid DNA Preparation and Isolation

2.4.1 Human Chromosome 8 Cosmid Library

A human chromosome 8 specific cosmid library, LA08NC01, representing 4 genome equivalents, was constructed at Los Alamos National Laboratory, New Mexico (Wood *et al.*, 1992). Chromosome 8 was isolated from a human x hamster hybrid cell line (UV20HL21-27), containing human chromosomes 4, 8, and 21 (Fuscoe *et al.*, 1986), by fluorescence-activated flow sorting (Deaven *et al.*, 1986). The chromosome 8 DNA was extracted, subjected to partial digestion by Sau3A1, dephosphorylated and ligated into *Bam*HI dephosphorylated sCos-1 vector (Evans *et al.*, 1989). The DNA was transfected into *E.coli* DHαMCR cells and plated onto agar plates containing kanamycin (50 μg/ml). Kanamycin resistant *E.coli* colonies were transferred into 96-well microtitre plates containing L broth and grown overnight at 37° C. Glycerol was added to a final concentration of 40% (Rose *et al.*, 1990). Four genome equivalents, represented

by 20,160 colonies, were arranged into 210 microtitre plates and stored at -70° C. Hybridization with both human and Chinese hamster DNA showed 85% human specificity. Nine loci mapping to different areas of chromosome 8 were used to screen the library. In all cases isolation of 1-4 cosmids was achieved (Wood *et al.*, 1992).

Agar Plates
4 g bactopeptone
2 g yeast extract
2 g NaCl
0.4 g Dextrose
4.8 g agar
400 ml dH_2O

2.4.2 Preparation of High Density Chromosome 8 Cosmid Library Filters

High density cosmid library filters were prepared by Ashley Howard using the robotic Biomek 1000 (Beckman Instruments Inc.). Hybond N+ membranes (Amersham) overlaid on agar plates containing kanamycin (50 μg/ml) were inoculated, by staggering, so that the clones from sixteen 96 well plates could be placed on a single filter (1,536 clones). Inoculated filters were grown overnight at 37° C. The DNA was lysed, denatured, and fixed to the membrane by the protocol of Grunstein and Hogness (1975). This high density procedure arrays the entire library on thirteen filters labeled A through M.

2.4.3 Isolation of Cosmid DNA

Cosmids were streaked, from glycerol stocks stored at -70° C, onto agar plates containing kanamycin (50μg/ml) and grown overnight at 37° C. Individual colonies were inoculated into 5 ml L broth and grown overnight, with shaking, at 37° C. Cosmid DNA was prepared by the methods of Birnboim and Doly (1979) and Ish-Horowicz and Burke (1981), with some

modifications. 1.5 ml broth was centrifuged in 1.5 ml eppendorf tubes at 13,000 rpm for 30 seconds to pellet the *E. coli* cells. The supernatant was completely removed, the pellet carefully resuspended in 100 µl ice cold Solution I and left for 3-5 minutes at room temperature. 200 µl of freshly made Solution II was added, mixed by inversion, and the tube placed on ice for 3-5 minutes. 150 µl of ice cold Solution III was added, mixed by inversion, and the tube was replaced on ice for 3-5 minutes. Samples were centrifuged for 5 minutes at 13,000 rpm, and 400 µl of solution was removed to a clean 1.5 ml eppendorf tube to which 800 µl of ice cold 95% ethanol or 400 µl room temperature isopropanol was added to precipitate the DNA. Samples were centrifuged at 13,000 rpm for 7-15 minutes, the alcohol removed and the pellet washed with 1ml ice cold 70% ethanol. Pellets were air dried and resuspended in 50 µl of TE(pH8.0)/RNase(25 µg/100 µl) followed by incubation at 37° C for 30 minutes to digest any RNA contamination.

1 µl of sample in 5 µl of stop buffer was run into 0.8% agarose, by electrophoresis, to determine the DNA concentration.

L Broth
4 g bactopeptone (tryptone)
2 g yeast extract
2 g NaCl
0.4 g D-glucose
400 ml dH_2O

Solution I
50mM glucose
25 mM Tris Cl (pH 8.0)
10 mM EDTA (pH 8.0)
dH_2O

Solution II
0.2 N NaOH
1% SDS

dH_2O

Solution III
3 M potassium acetate
5 M acetic acid
dH_2O

1X TE
10 mM Tris-Cl (pH 8.0)
1 mM EDTA

2.5 Bacterial Artificial Chromosome (BAC) DNA Isolation and Preparation

2.5.1 Human Chromosome 8p Enriched BAC library

A human chromosome 8p enriched BAC library was prepared by Michael Schertzer. Inter-*Alu* PCR allows the amplification of DNA between two *Alu* repeats separated by a distance compatible with PCR amplification (Nelson *et al.*, 1989). The cell line 1HL12 (Wagner *et al.*, 1991) retains only the short arm of human chromosome 8. Inter-*Alu* PCR using the ALE-1 and ALE-3 primers (table 2) (Cole *et al.*, 1991) was performed on this cell line using standard PCR conditions (see section 2.1). The PCR products were used as hybridization probes (see sections 2.12, 2.13) (Sambrook *et al.*, 1989) to high density filters of the Research Genetics BAC library consisting of 27,698 clones spotted in duplicate (55,296 total clones) onto 5 high density filters. Strong positives were picked and inoculated into 34 - 96 well plates (3,237 clones in total) containing L broth. Clones were grown overnight at 37° C. Glycerol was added to a final concentration of 40% (Rose *et al.*, 1990) and clones were frozen at -70° C.

2.5.2 Human Chromosome 8p Enriched High Density Filters.

High density BAC library filters were prepared by Michael Schertzer as in section 2.4.2 for cosmid high density filters, with the exception that chloramphenicol (37.5μg/ml) was used in the place of kanamycin. This high density procedure arrays the entire library (3,237 clones) on 2 filters labeled A and B.

2.5.3 Isolation of BAC DNA

Bacterial artificial chromosomes were streaked from stocks stored at -70° C onto agar plates containing chloramphenicol (37.5 μg/ml) and grown overnight at 37° C. Individual colonies were inoculated into L broth containing chloramphenicol (30 ng/ml) and grown overnight, with shaking, at 37°C. BACs were isolated by alkaline lysis, as in section 2.4.3, with the following modifications. After the addition of Solution III and centrifugation for 10 minutes the supernatant was carefully poured into a clean 1.5 ml eppendorf to avoid shearing of the DNA. An equal volume of room temperature isopropanol was added and the DNA was pelleted by centrifugation at room temperature for 15 minutes. DNA was washed with 70% ethanol, dried, and 45 μl of 1X TE(pH8.0)/RNase(25 μg/100 μl) was added. The pellet was carefully scraped from the side of the eppendorf tube and allowed to dissolve at room temperature for 20-30 minutes. The bottom of the eppendorf tube was gently tapped to resuspend the DNA, briefly centrifuged, and placed at 37° C for 45 minutes to allow any contaminating RNA to be digested. Samples were stored at -20° C.

2.6 Yeast Artificial Chromosome (YAC) DNA Preparation and Isolation

2.6.1 Centre d'Étude du Polymorphisme Humain (CEPH) Mega YACs

A chromosome 8 enriched subset of YACs from the CEPH Mega-YAC library (Bellanne-Chantelot *et al.*, 1992) was isolated by hybridization of YAC Inter-*Alu* PCR products onto a chromosome 8 specific somatic cell hybrid panel (Chumakov *et al.*, 1992). Yeast lines containing YACs found positive for chromosome 8p markers, summarized at the MIT/Whitehead Institute database (http://www.genome.wit.mit.edu), were individually purchased from Research Genetics, Inc.

2.6.2 Isolation of YAC DNA

Cells containing YACs were streaked from glycerol stocks stored at -70° C (Rose *et al.*, 1990) onto AHC minimal media agar plates (Brownstein *et al.*, 1989) and allowed to grow for 3 days at 30° C.

2.6.2.1 Protocol 1: YAC DNA Preparation Using Glass Beads

A single red colony was inoculated into 5 mls of YPD in a 15 ml tube and grown for 24 hours, with shaking, at 30° C. Since yeast cells sediment at the bottom of the 15 ml tube, 4 mls of YPD were drawn off and the remaining 1 ml, containing the majority of the yeast cells, was placed in a 1.5 ml eppendorf tube. Cells were pelleted by centrifugation, washed with 500 µl distilled water, and resuspended in 500 µl GDIS. The cells were added to 1.5 ml eppendorf tubes containing 0.35 g of acid washed glass beads and 200 µl of 25:24:1 phenol:chloroform:isoamyl alcohol. Samples were vortexed for 2 ½ minutes, 200 µl of distilled water was added, mixed, and the samples were centrifuged at 13,000 rpm for 2 minutes. The upper aqueous layer, containing the

DNA, was removed to a new 1.5 ml eppendorf tube to which an equal volume of room temperature isopropanol was added to precipitate DNA. DNA was collected by centrifugation at 13,000 rpm for 15 minutes. Pellets were washed with 1ml 70% ethanol, dried, and resuspended in 50 µl TE.

GDIS
2% Triton X-100
1% SDS
100mM NaCl
10mM Tris-Cl pH 8.0
1mM EDTA

Acid washed glass beads
Approximately 10 g of glass beads were washed with 5 ml of 0.1N HCl, rinsed three times with 5 ml of distilled water and excess water was removed with a pipetman. Beads were autoclaved for 20 minutes and dried.

2.6.2.2 Protocol 2: Minipreparation of YAC DNA

A single red colony was inoculated into 5 ml of YPD containing ampicillin (100ng/ml) in a 15 ml tube and allowed to grow for 2 days at 30°C, with shaking. 1.5 ml of the sample was centrifuged in a 1.5 ml eppendorf tube at 13,000 rpm for 2 minutes. The supernatant was completely removed, the pellet resuspended in 240 µl of lysis solution and incubated at 37° C for 1 hour. The sample was pelleted by centrifugation at 13,000 rpm for 2 minutes, the supernatant removed, the pellet resuspended in a mixture of 100 µl YTE plus 10 µl 10% SDS and incubated at 65° C for 20 minutes. 40 µl of alkaline lysis solution III was added, mixed, and the sample was placed on ice for 30 minutes. The cellular debris was separated from the DNA by centrifugation for 3 minutes at 13,000 rpm. 150 µl of the supernatant was removed to a fresh 1.5 ml eppendorf tube and two volumes of 95% ethanol were added to precipitate the DNA. The

DNA was pelleted by a 15 minute centrifugation at 13,000 rpm, washed with 70% ethanol, dried, resuspended in 100 µl of 1X TE/RNase(25µg/100 µl), and incubated for 1 hour at 37° C to degrade any contaminating RNA.

Sorbitol Solution
0.9 M Sorbitol
0.1 M ethylene diamine tetra-acetic acid (EDTA), pH8.0
100 mM tris (hydroxymethyl) aminomethane (Tris-Cl), pH8.0

Lysis Solution
200 µl sorbitol solution
20 µl 1:25 sorbitol solution:β-mercaptoethanol
20 µl sorbitol solution containing a few grains of lyticase or zymolase

YTE Solution
50 mM Tris-Cl, pH8.0
20 mM EDTA, pH8.0

AHC Minimal Media (ura-, trp-) Plates
0.67 g yeast nitrogen base w/o amino acids
1 g acid hydrolysed casein
2 g dextrose
1.86 mg adenine hemisulfate or adenine hydrochloride
1.2 g agar
100 ml dH_2O
Adjust to pH 5.8 with concentrated HCl (7 µl)

YPD Media
4 g tryptone (peptone)
2 g yeast extract
4 g dextrose
200 ml dH_2O

2.7 Restriction Enzyme Digestion

2.7.1 Standard Restriction Enzyme Digestion

Digestion of approximately 1 µg of DNA was carried out in either 20 µl or 25 µl final volumes, in 1.5 ml eppendorf tubes. Final concentrations of: 1X Enzyme React buffer, 1X BSA

and 4U of enzyme, were used. Digestion was carried out for 1½ hours at 37° C, except for *Bss*HII which was incubated at 50° C. Digestion was stopped by the addition of 5 µl of stop buffer.

2.7.2 Double Restriction Enzyme Digestion

When DNA was digested by two enzymes, equal amounts of each enzyme were used, and the most appropriate buffer to allow complete digestion of the DNA was chosen. When the enzymes required incompatible buffers for optimal digestion, the DNA was digested using standard conditions with one enzyme. After completion of the first digestion, the total volume was increased to 100 µl with distilled water, an equal volume of TE-equilibrated phenol was added, and the solution mixed by vortexing. Samples were centrifuged for 2 minutes, the upper aqueous layer removed to a clean eppendorf tube, an equal volume of Sevag's Solution added, vortexed, centrifuged, and the upper aqueous layer removed to a clean eppendorf tube. A 1/10 volume of 3M ammonium acetate was added, 2 volumes of 95% ethanol were added and the DNA was allowed to precipitate at -70° C for 30 minutes. Samples were centrifuged for 15 minutes at 13,000 rpm to pellet DNA, washed with 70% ethanol, dried, resuspended in 16 µl 1X TE and digested with the second restriction enzyme. Digestion was stopped by the addition of 5 µl of stop buffer.

Sevag's Solution
24:1 chloroform/isoamyl alcohol

10X BSA
bovine serum albumin fraction V 1mg/ml

Restriction Enzymes and buffers
*Eco*RI(Gibco/BRL), 10X REACT®3
*Sst*I (Gibco/BRL), 10X REACT®2

*Bgl*II (Gibco/BRL), 10X REACT®3
*Bss*HII (NEB), 10X NEB Buffer for *Bss*HII digests
*Bam*HI (Gibco/ BRL), 10X REACT®3
*Hae*III (Gibco/BRL), 10X REACT®2
*Kpn*I (Gibco/BRL), 10X REACT®4
*Not*I (NEB), 10X NEB3
*Nde*I (Gibco/BRL), 10X REACT®3
*Eco*RI/*Pvu*I(Gibco/BRL), 10X REACT®7
*Hind*III (Gibco/BRL), 10X REACT®2
*Sac*II (Gibco/BRL), 10X REACT®2

2.8 Agarose Gel Electrophoresis

2.8.1 Standard Gel Electrophoresis

Agarose gels were prepared by mixing appropriate amounts of agarose (Sigma) with 1X TBE to achieve the desired concentration of gel (w/v). The agarose was dissolved by heating in a microwave oven, cooled until warm, ethidium bromide (EtBr) was added to a final concentration of 0.5 μg/ml, the gel was poured in casting trays and allowed to solidify. Electrophoresis was carried out in 1X TBE. Digested DNA was run into agarose gels of the appropriate concentration based on expected fragment sizes. DNA digested with enzymes which cut infrequently, *Bss*HII, *Kpn*I, and *Nde*I, were run into 0.4% agarose stabilized by a bottom layer of 1% agarose gel containing no EtBr. *Sst*I digests were run into 0.66% agarose gels, and digests using frequently cutting enzymes, *Eco*RI, *Eco*RI/*Pvu*I, *Bam*HI, and *Bgl*II, were run into 0.8% agarose gels. In all cases 5 μg λ DNA digested with *Hind*III and *Sac*II was run alongside DNA samples as a molecular weight marker. Agarose gels containing restriction enzyme digested DNA were run overnight at 30 volts.

PCR products were run into 2-3% agarose gels prepared in the above manner. Samples were run into agarose gels for 2-4 hours at 80-150 V in order to achieve the desired separation. In all cases a molecular weight marker of 2.5 μg ϕX174 DNA digested with *Hae*III (Gibco/BRL) was run alongside PCR samples.

10X TBE
54 g Tris base
27.5 g boric acid
10mM EDTA, pH8.0
dH_2O to 500 ml

λ DNA marker
40 μl λ DNA (500 ng/μl)
16 μl 10X BSA
16 μl 10X REACT®2
88 μl Milli Q filtered dH_2O
5 μl *Hind*III
5 μl *Sst*II/*Sac*II
40 μl of stop buffer added after 1 ½ hours at 37° C

2.8.2 Pulsed Field Gel Electrophoresis (PFGE)

A hexagonal array pulsed field gel apparatus was used. Three litres of 1X TBE were prepared from autoclaved 10X TBE. 150 ml of this solution was removed to make a standard 1% agarose gel containing no EtBr. Approximately 2 litres of the remaining 1X TBE were poured into the PFG apparatus, the pump and cooler were turned on, and the 1X TBE was cooled to 12°C. The 1% agarose gel, prepared as in section 2.8.1, was cast in the PFG casting tray and allowed to cool. The comb was removed and a small slice of Low Range PFG marker(NEB) or Lambda ladder PFG marker (NEB) in agarose was loaded into the two lanes that would flank the samples. The casting tray was placed into the PFG apparatus, the volume of 1X TBE was adjusted with the remaining solution to just cover the gel. Samples were digested with *Not*I and loaded into the

gel. The run conditions were as follows: 36 second pulses alternating from North/South to East/West, at 200 V for 20 hours. The gel was removed from the casting tray and stained in a 0.5 μg /ml EtBr solution for 30 minutes.

2.9 DNA Imaging for Agarose Gels

Agarose gels were placed on a FOTO/Convertible (FOTODYNE) Ultraviolet radiation box to visualize the DNA containing intercalated EtBr. The gels were photographed using the Scion Gel Imaging System via a CCD camera linked to a Power MacIntosh computer.

2.10 Southern Transfer

Agarose gels were photographed alongside a ruler, trimmed to remove excess agarose and blotted onto nylon membranes by the method of Southern (Southern, 1975). Those gels containing DNA fragments of 15 kb or more were first treated with 0.25 N Hydrochloric Acid for 30 minutes to depurinate the DNA. Gels were placed in 1.5 M NaCl/0.5 M NaOH for 30 minutes to denature the DNA. The solution was removed and replaced with neutralization solution (1 M Tris/1.5 M NaCl) for 30 minutes, followed by rinsing with distilled water. A Southern blot apparatus was assembled. Approximately 500 ml of 10X SSC was placed in a pyrex dish, a piece of glass was placed across the top of the dish and two sheets of 3M Whatmann paper wet with 10X SSC were placed on the glass to form a wick. Air bubbles were removed by rolling a pipet over the wick. The agarose gel was placed on the wick, a piece of Hybond N+ nylon membrane (Amersham) cut to the gel dimensions was briefly wet with distilled water and placed over the gel, ensuring that the top of the membrane was flush with the wells of the agarose gel. A piece of Whatmann paper cut to the gel dimensions was wet with

distilled water and placed on top of the membrane, followed by a similar piece of dry Whatmann paper. The area surrounding the gel was covered with plastic and paper towel was placed over the gel to draw the solution up through the gel and allow the DNA to bind to the charged nylon membrane. The apparatus was left overnight, disassembled the next morning, the membrane placed between two pieces of dry Whatmann paper and placed at 80° C for 2 hours to cross link the DNA to the membrane.

20X SSC
175.3 g NaCl
88.2 g sodium citrate
dH_2O to 1 litre

2.11 Radioactive Labeling of DNA

2.11.1 Random Primer Labeling

The random prime labeling method of Feinberg and Vogelstein (1984 a, b) was used. 15 µl of 1 ng/µl probe was boiled for 5 minutes, placed on ice for 5 minutes and mixed with 2.5 µl α-$dATP^{32}$ (NEN Dupont 10 µCi/µl), 5 µl 5X Oligolabeling buffer without dATP (OLB-A) containing random hexamer DNA, 2.5 µl 10X BSA, and 1U Klenow enzyme (Pharmacia or Gibco/BRL). Labeling was allowed to occur at room temperature overnight.

5X OLB-A
Mix solutions A:B:C to a ration of 100:250:150
Solution O: 1.25 M Tris-Cl pH8.0
0.125 M $MgCl_2$
Solution A: 1 µl Solution O
18 µl β-mercaptoethanol
5 µl each dGTP, dCTP, dTTP
(each dNTP is 0.1 M in 3 mM Tris pH7.0, 0.2 mM EDTA)
Solution B: 2 M Hepes pH6.6
Solution C: Hexadeoxyribonucleotides suspended in 1X TE to 90 OD/ml

2.11.2 Klenow Primer Extension Labeling

15 µl of 1 ng/µl probe plus 0.5 µl of each primer pair was boiled for 5 minutes, placed on ice for 5 minutes, and mixed with 1µl α-dATP^{32} (NEN Dupont 10uCi/µl), 5 µl 5X KEL buffer, 2.5 µl 10X BSA, 1U Klenow enzyme (Pharmacia or Gibco/BRL). Samples were placed at 37° C for 1 ½ hours to allow labeling by Klenow primer extension to occur.

5X KEL buffer
OLB-A with the exception that Solution C is dH_2O

2.11.3 γ-end Labeling

DNA was labeled in a 10 µl reaction containing 50 ρmoles of DNA primer, 1X Kinase Forward Reaction Buffer (Gibco/BRL), 1U T4 Kinase (Gibco/BRL) and 25 µCi γ-dATP. Labeling was carried out at 37° C for 60 minutes followed by 5 minutes at 95° C to heat inactivate the enzyme. The primer was stored at -20° C.

2.11.4 Nick Translation of Plasmid DNA

1 µg of plasmid DNA (usually pBluescript) was labeled by nick translation using the Gibco/BRL nick translation kit. To 1 µg of DNA was added 5 µl Solution A1, 5 µl ^{35}S-αATP (NEN/Dupont), and dH_2O to 45 µl. 5 µl of Solution C (DNA Polymerase/DNaseI) was added and labeling was allowed to proceed for 1 hour at 14° C. 5 µl of Solution D (300 mM Na_2EDTA, pH8.0) was added to stop the reaction. The labeled plasmid was diluted with dH_2O to a final concentration of 5 ng/µl.

Solution A1
0.2 mM dCTP, dGTP, dTTP
500 mM Tris-HCl, pH7.8
50 mM $MgCl_2$
100 mM 2-mercaptoethanol

2.12 Pre-Hybridization Treatment of Probe

2.12.1 Removal of Unincorporated Nucleotides

25 µl of nick translation stop buffer (NTSB) was added to the probe upon completion of labeling. The probe was centrifuged at 1,000 rpm through a Sephadex G-25 spin column. The Sephadex G-25 beads capture unincorporated nucleotides while allowing nucleotides incorporated into DNA to remain in solution and pass through the column. The solution containing the probe was brought back up to 50 µl by the addition of Milli Q filtered distilled water.

NTSB
50mM EDTA
20mM NaCl
0.1% SDS
500 mg/ml salmon sperm DNA

2.12.2 High Complexity Probes

Probes containing little or no common repetitive DNA were boiled for 5 minutes, placed on ice for 5 minutes and used for hybridization. These included PCR products, the megasatellite DNA and the subcloned fragments of the cosmids 153G8 and 39A7

2.12.3 Low Complexity Probes

Probes containing common repetitive DNA were mixed with 5.5 µl 25X SSC and 15 µl human placental DNA (10.6 mg/ml) (Sigma), boiled for 5 minutes to denature the DNA and allowed to preanneal at 65° C for 1 hour before being used for hybridization. These probes included BAC inserts and end sequences known to contain repetitive DNA.

2.12.4 Probes to High Density Filters

Probes being used to screen high density filters were mixed with 10 µl of plasmid (5 ng/µl) labeled with ^{35}S-αATP by nick translation, to allow background visualization of high density filter grid, boiled for 5 minutes to denature the DNA, placed on ice for 5 minutes and used for hybridization.

2.13 Hybridization, Post-Hybridization Washes, Autoradiography

Hybridizations were carried out in heat-sealable hybridization bags containing the probe and enough hybridization buffer to wet the filter plus 5 ml. Hybridization was allowed to occur overnight at 65° C followed by a 5 minute wash in 1X SSC/1% SDS at room temperature and a 50 minute stringent wash in 0.2X SSC/1% SDS at 65° C. Filters were partially air dried, wrapped in plastic wrap and placed on film on a lightening plus screen at -70° C or at room temperature for the appropriate length of time required for a clear signal. When screening libraries, positive signals were read off the high density filters and the grid locations were matched with the corresponding clone address. If necessary, blots were stripped of signal by pouring boiling 0.5% SDS over them and allowing the solution to cool to room temperature, rinsing with distilled water and allowing them to dry.

Hybridization Solution
150 ml 20X SSC
50 ml 5X Denhardt's Solution (filter sterilized)
7.5 ml 20% SDS
50 ml 100 mg/ml salmon sperm DNA
242.5 ml distilled water

100X Denhardt's Solution
4% Ficoll 400
4% polyvinyl pyrrolidone
4% (w/v) bovine serum albumin (BSA)

2.14 Extraction of DNA from Agarose Gel

2.14.1 Qiagen Gel Extraction

In order to subclone restriction enzyme fragments and PCR products the DNA must be free of agarose contaminants. The QIAEXII Agarose Gel Extraction Kit (Qiagen) was used to extract DNA that was to be used for subcloning experiments. The DNA band was excised from the agarose gel and placed in an eppendorf tube containing Buffer QX1, which solubilizes agarose and creates a high salt environment, and QIAEX II silica gel particles, which adsorb nucleic acids in the presence of high salt. The solution was incubated at 50° C for 10 minutes with intermittent vortexing to keep the QIAEX II in suspension. The tube was centrifuged at 13,000 rpm for 30 seconds and the supernatant removed. The pellet was washed once with Buffer QXI to remove residual agarose and twice with Buffer PE to remove salt contamination. The pellet was air-dried followed by a 5 minute incubation in 20 µl distilled water for elution of the DNA from the QIAEXII particles occurs. The QIAEXII particles were pelleted by centrifugation and the supernatant, containing the DNA, was removed to a clean eppendorf tube and stored at -20° C.

2.14.2 Low Melting Point Gel Extraction

The band of interest was excised from the agarose gel and placed in the well of a low melting point (LMP) agarose gel (Sigma). The DNA was run into the LMP gel in fresh 1X TBE at 80 V until the band had migrated approximately 5 cm into the gel. This removes contaminating

plasmid DNA. The band was extracted, placed in 1X TE to a concentration of 1 ng/μl, boiled for 5 minutes to liberate the DNA from the agarose and stored at 4° C.

2.14.3 DEAE Protocol

A cut in the agarose gel was made just below the band of interest. The gel was pried apart using clean forceps and a small piece of DEAE paper (Schleicher & Schuell) was inserted. DEAE paper carries a functional group diethylaminoethyl in its protonated form that has a binding capacity of 20 $\mu g/cm^2$ for DNA, RNA, and protein. The gel was run at 100 V for approximately 5 minutes until the band had just contacted the DEAE paper. The DEAE paper was removed and placed in an 1.5 ml eppendorf tube containing 500 μl of NaCl, creating a high salt environment. The band was eluted from the DEAE paper at 65° C for 45 minutes. The paper was checked under UV illumination for any residual DNA, removed, and two volumes of 95% ethanol were added the tube. The sample was placed at -70° C, to facilitate precipitation, until frozen. The DNA was collected by centrifugation at 13,000 rpm for 15 minutes. The pellet was washed in 1 ml 70% ethanol, air dried and resuspended in the appropriate amount of 1X TE.

2.15 Subcloning of Cosmid Fragments

Cosmid fragments were subcloned into either pBluescriptII KS (Stratagene) vector or a modified pBluescript vector. The modified vector has the SpeI and XbaI restriction sites replaced with an AscI restriction site allowing *Bss*HII fragments to be cloned (DeBella et al., unpublished, 1998). Two selectable markers are contained in pBluescript: the ampicillin resistance gene, which allows selectable growth on ampicillin of those clones containing the vector, and the coding region for the amino-terminal region of the β-galactosidase gene (lacZ).

Contained within the cloning region of the lacZ gene is a polycloning site which does not disrupt the reading frame, and therefore the gene product, unless a DNA fragment has been cloned into it. When the vector is present within a host such as DH5α *Escherichia coli*, which encodes the carboxy-terminal portion of the β-galactosidase gene, the two fragments associate to form an enzymatically active protein (Ullman et al., 1967). Clones containing the active protein appear blue when grown in the presence of 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) and isopropyl-thiogalactoside (IPTG). Insertion of DNA into the polycloning site of the vector, disrupting the transcription of the amino-terminal portion of the gene, results in white colonies.

2.15.1 Preparation of pBluescript Vector

The pBluescriptII KS or the pBluescriptII KSAsc was cut with the appropriate restriction enzyme to achieve compatible ends with the DNA fragment to be subcloned. The vector was digested according to standard restriction enzyme conditions in a final volume of 50 µl. The vector DNA was precipitated with 95% ethanol, washed with 70% ethanol, air-dried and resuspended to a final concentration of 100 ng/µl. 1 µl was run out on an agarose gel to ensure complete digestion. The vector was stored at -20° C.

2.15.2 Subcloning

Cosmid *Eco*RI fragments were randomly cloned in order to isolate fragments for study. 1 µg of DNA was double digested with *Eco*RI/*Pvu*I using the double restriction enzyme digestion procedure (section 2.7.2). 5 µl of stop buffer was added and the sample was run into a 0.8% gel in 1X TBE at 30 V, overnight. *Pvu*I cleaves once within the vector and it was determined if there were any sites of cleavage in the cosmid insert. Once this had been determined, 1 µg of

DNA was digested with *Eco*RI/*Pvu*I, precipitated, and resuspended in 16 µl dH_2O. By digesting with *Eco*RI in the presence of *Pvu*I, the vector is cut into two pieces incompatible with ligation into *Eco*RI cleaved pBluescript. This selects against the ligation of vector DNA into pBluescript.

2.15.3 Ligation of Restriction Enzyme Digested Fragments into pBluescript

Ligation reactions were carried out in a final volume of 20 µl containing approximately 100 ng QIAGEN gel purified DNA or 1 µg restriction enzyme digested cosmid DNA, 1X NEB T4 ligase buffer, 20 ng pBluescript digested with the appropriate restriction enzyme, and 1U T4 DNA ligase (NEB). A positive control, containing only pBluescript with cohesive ends was included with each ligation performed. The ligation reaction was carried out overnight at 14° C.

2.15.4 Deletion Clones

Plasmid deletion clones were constructed by digestion of 1 µg of the original clone with an enzyme present within the polylinker and the insert, followed by religation using the procedure outlined in section 2.15.3, with the exception that pBluescript was not included in the reaction.

2.16 Transformation of Ligated DNA into Competent *E.coli* Cells

Competent DH5α *E. coli* cells (Inoue et al., 1990) were removed from -70° C and thawed on ice. 50 µl was placed into a precooled 15 ml tube and 10 µl of the ligation reaction was added, mixed, and left on ice for 30 minutes. The cells were heat shocked for 45 seconds at 42° C and returned to ice for 2 minutes. 400 µl L broth was added and the cells were allowed to divide for 45 minutes at 37° C. 120 µl of transformed cells was pipeted onto agar plates, spread, and

incubated overnight at 37° C. Agar plates containing ampicillin (50 μg/μl) were used for cosmid ends and XIA plates were used for DNA subcloned into pBluescript and for deletion clones.

XIA plates:
Agar plates (see section 2.4.1)
plus 20 mg ampicillin, 25 mg X-gal, 50 mg IPTG
NOTE: plates must be cooled and stored in the dark

2.17 DNA Sequencing

2.17.1 DNA Preparation for Sequencing

In order to obtain accurate sequencing results, the DNA to be sequenced must be especially pure. This is achieved through use of the Qiagen Plasmid Mini Kit (QIAGEN). A modified alkaline lysis protocol is used to isolate DNA: 1.5 ml of L broth, inoculated with a single bacterial colony containing the subcloned fragment of interest, grown overnight, was centrifuged and the pellet resuspended in 300 μl of cold Buffer P1. 300 μl of ice cold Buffer P2 was added, mixed gently by inversion and incubated on ice for 5 minutes. The sample was centrifuged for 10 minutes to pellet cellular debris, and the supernatant removed to a QIAGEN-tip pre-equilibrated with 1 ml of Buffer QBT. Buffer QBT and the supernatant create a low salt environment, allowing DNA to bind to the anion-exchange resin contained within the QIAGEN-tip. Once the supernatant had drained through the tip, 4 ml of Buffer QC, a medium concentration salt solution, was added to the tip to remove impurities, such as RNA and proteins. The purified DNA was eluted into a clean 1.5 ml eppendorf tube by addition of 800 μl of Buffer QF, a high salt buffer at pH7.0. The DNA was precipitated using 0.7X volume isopropanol, pelleted by centrifugation at room temperature, and washed with 70% ethanol to remove excess salt. The pellet was resuspended in 15 μl Milli-Q purified dH_2O. 1 μl of each sample was

digested with *Eco*RI (section 2.7.1) and the fragments were separated on an agarose gel to determine DNA concentration.

2.17.2 Automated Sequencing

Sequencing was carried out by the automated sequencer (Applied Biosystems Model 373 Stretch), using either the M13F (5′-CCCAGTCACGACGTTGTAAAACG-3′) or the M13R (5′-AGCGGATAACAATTTCACACAGG-3′) primer, which flank the cloning sites of the pBluescript vector. Sequence was obtained from 500 ng of QIAGEN purified double stranded DNA template using ABI's AmpliTaq FS DyeDeoxy™ Terminator Cycle Sequencing. All four base reactions are carried out in one eppendorf tube on a thermocycler. The sequence obtained by this method can be 98.0% accurate to more than 650 base pairs. Sequencing was carried out by the U.B.C. sequencing laboratory. The procedure used is as follows: 8.0 µl of terminator premix, 500 ng of template, 3.2 pmol of primer (M13 or T3), and dH_2O to a final volume of 20.0 µl were mixed in a 0.6 ml eppendorf tube. The reaction mixture was overlaid with a drop of mineral oil and the tubes were placed in a preheated (96° C) thermocycler. Twenty-five cycles of: 96° C for 30 seconds, 50° C for 15 seconds, and 60° C for 4 minutes, were carried out followed by incubation at 4° C until reaction was removed from the thermocycler. Each reaction was transferred to a 1.5 ml eppendorf and precipitated with 1/10 volume 3 M sodium acetate, pH4.6, and 2 volumes 95% ethanol. This precipitation step removes excess dye terminators. The DNA was pelleted by centrifugation, dried, and stored at -20° C. Fragments were separated on sequencing gels using an automated sequencer (Applied Biosystems Model 373 Stretch).

2.17.3 DNA Sequence Analysis

2.17.3.1 BLAST

Sequences obtained from the subcloned fragments were analyzed using the basic local alignment search tool (BLAST). BLAST uses a rapid database searching algorithm that optimizes local similarities between sequences and extends these alignments based on defined match and mismatch criteria (Altschul *et al.*, 1990). The statistical significance of any similar segments between the query sequence and a given database sequence is evaluated and reported. Only P values of 1.1×10^{-21} or less were examined. The P value is the likelihood of the match representing a random alignment. For the subclones analyzed, any match to the megasatellite sequence was at a P value less than 2×10^{-33}. BLAST queries Genbank, the National Institute of Health database, maintained at the National Center for Technology (NCBI) (http://www.ncbi.nlm.nih.gov), which contains all reported nucleotide and protein sequences. BLASTnr accesses the non-redundant database, BLASTsts accesses the STS database and BLASTest accesses the EST database for matching sequences. All three types of BLAST queries were carried out on the sequenced DNA.

2.17.3.2 Clustal W

Sequences obtained from the subcloned fragments of 153G8 and 39A7 were compared to each other, and to the 4p megasatellite sequence (Genbank accession #: D38378) using the Clustal W program available at http://www2.ebi.ac.uk/clustalw (Thompson *et al.*, 1994). Clustal W aligns sequences with respect to one another, allowing for insertions and deletions. Problems occur when two sequences are of very different lengths, or when there are long regions which can not

be aligned. Therefore, when comparing subclone sequence with megasatellite sequence, BLAST alignments were consulted for regions of homology, the regions were aligned using Clustal W, followed by alignment of sequences 'by eye'.

2.18 Cosmid Termini Isolation

2.18.1 Isolation of T3 End

The following procedure is diagrammed in figure 6. *Bgl*II is a 6 cutter that cleaves fairly frequently in human DNA. s-Cos1 vector contains a single *Bgl*II restriction enzyme site. Cosmids were digested with *Bgl*II using standard conditions (section 2.7.1), the DNA was precipitated, and religated (section 2.15.3 with the exception of dH_2O used in place of plasmid DNA). Religation results in a variety of products including deleted cosmids made up of the T3 end of the cosmid insert and 4300 bp of vector DNA containing the ampicillin resistance gene. Transformation (section 2.16) of these products into competent *E. coli* cells and growth on agar plates containing ampicillin results in a selection for the T3 end of the cosmid insert.

2.18.2 Isolation of T7 End

1 μg of cosmid was digested with *Eco*RI using standard conditions and run overnight at 30V into a 0.8% agarose gel in 1X TBE (sections 2.7.1, 2.8.1). The gel was transferred to nylon membrane by Southern blotting, hybridized overnight at 42° C with T7 primer labeled by γ-end labeling, washed, placed on film (sections 2.10 , 2.11.3, 2.12, 2.13), and the size of the *Eco*RI fragment containing the T7 end was determined. The cosmid was again digested with *Eco*RI and run into 0.8% agarose in 1X TBE at 30 V, overnight. The band of interest was extracted by low melting point agarose gel extraction (section 2.14.2).

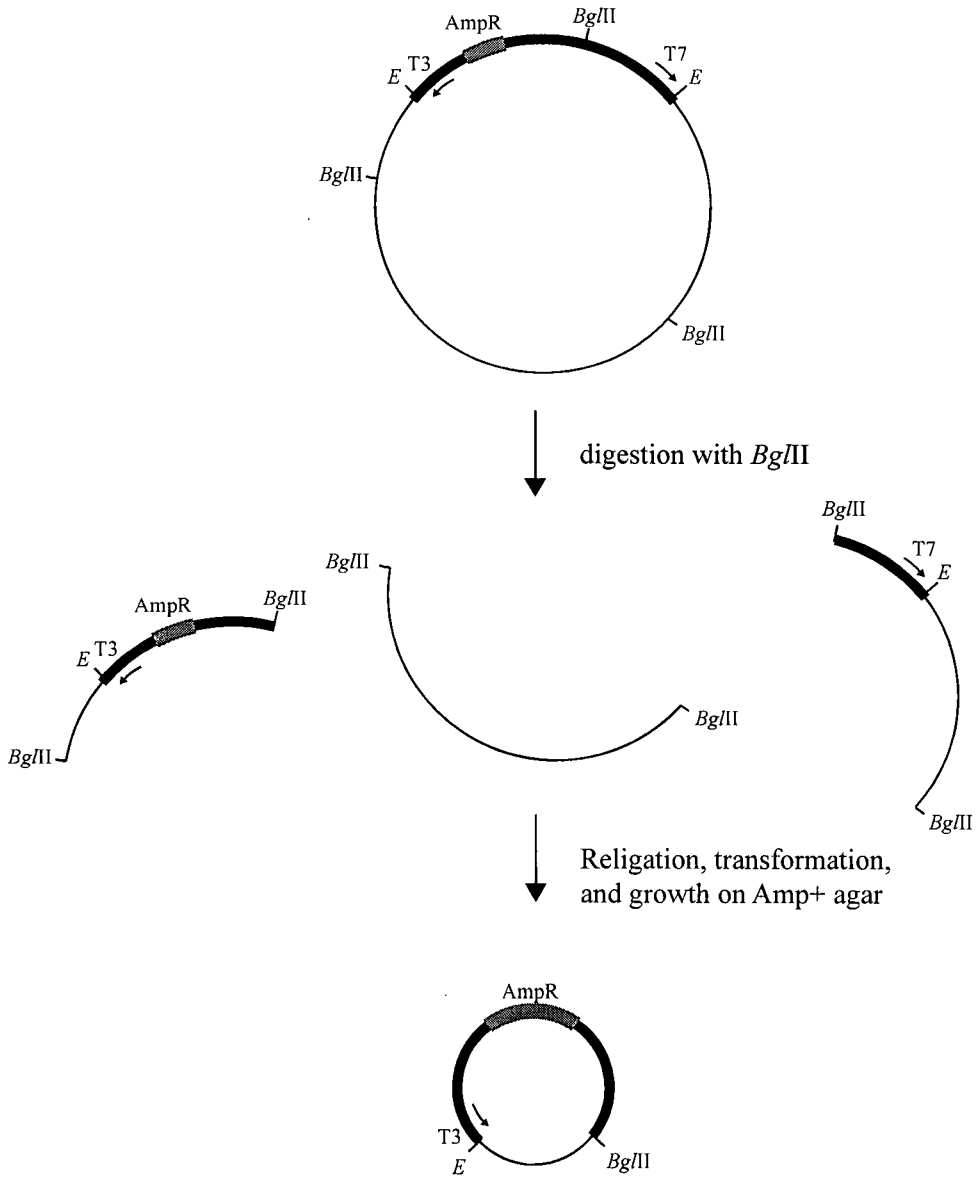


Figure 6: Construction of T3 end deletion clones from cosmid clones. Dark line represents vector DNA, light line represents human DNA, *E*=*Eco*RI. Selection for those colonies that grow on ampicillan results in isolation of T3 end of the human DNA insert.

2.19 BAC Termini Isolation by Bubble PCR

Bubble PCR was initially developed as a method for isolating YAC termini (Riley et al., 1990). We have modified the technique to allow isolation of BAC termini. Briefly, ligation of a linker containing a region of mismatch, to restriction enzyme digested DNA, allows PCR to be carried out preferentially from only those fragments that contain vector plus insert DNA (see figure 7).

2.19.1 Digestion of DNA

Approximately 500 ng of BAC DNA was digested with DdeI in a 30 µl reaction volume, final concentration of 1X BSA, 1X REACT®2, 4U enzyme. Reactions were carried out at 37° C for 2 hours. DNA was precipitated and resuspended in 12 µl of distilled water.

2.19.2 Preparation of Linker DNA

The bubble linker was prepared by mixing equimolar amounts of bottom strand oligo 221 (5′-CTCTCCCTTCTCGAATCGTAACCGTTCGTACGAGAATCGCTGTCCTCTCCTTG-3′) and top strand oligo 222 (5′-TNACAAGGAGAGGACGCTGTCTGTCGAAGGTAAGGAACGGA CGAGAGAAGGGAGAG-3′) with 25 mM NaCl. The solution was boiled for 2 minutes followed by incubation for 5 minutes at 65° C. The solution was left at room temperature to cool.

2.19.3 Ligation of Linker to BAC DNA

The digested DNA was mixed with 15 µl of linker DNA, 3 µl of 5X NEB ligation buffer and 1 µl of NEB T4 DNA ligase enzyme. The ligation mixture was left at 14° C overnight and adjusted with 1X TE to a final volume of 100 µl.

2.19.4 PCR Isolation of BAC Termini

10 µl of the ligation mixture was used in a standard PCR as in section 2.1 using 10 pmole each of oligo 224 and oligo T7 (T7 end) or oligo 224 and oligo SP6 (SP6 end) (see table 2 for primer sequences). The PCR annealing temperature was 40° C. Two samples of each BAC were run for T7/224 and 1 sample for SP6/224. 1 µl of each reaction was saved in case of insufficient amplification. One sample of T7/224 was digested with Pst1 to remove vector sequences. This was done by drawing out the sample from below the oil and placing it on parafilm. The drop of sample was rolled across the parafilm to remove any residual mineral oil and then placed in a 500 µl eppendorf tube. A standard restriction digest by Pst1 was carried out using the PCR sample as the DNA source (without any precipitation from the PCR mix). After 1 hour, stop buffer was added to the digested sample and the remaining samples, and they were loaded into a 2% agarose gel. Gels were run for approximately 2 hours at 100 V in 1X TBE. Digested T7 ends and undigested SP6 ends were extracted from the gel either by the DEAE protocol or by LMP agarose excision (section 2.14).

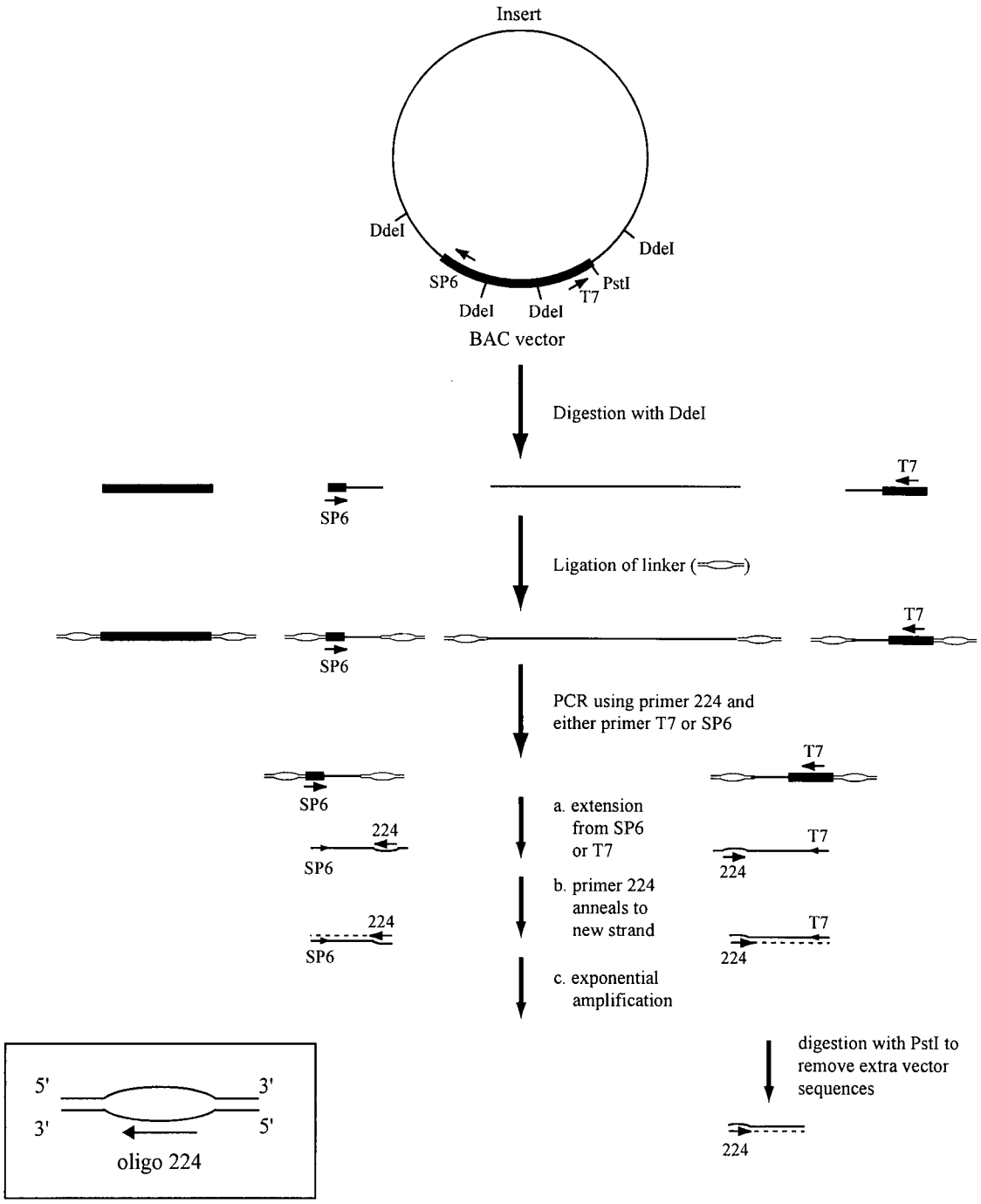


Figure 7: Bubble PCR protocol for BAC end isolation. Inset: The bubble linker. Oligo 224 is identical to one strand of the linker but not complementary to the other strand due to a region of mismatch. Therefore, amplification can not occur from oligo 224 unless extension has occurred from either the SP6 or T7 primer allowing the SP6 and T7 ends of a human DNA insert to be preferentially amplified.

2.20 Preparation of DNA from Whole Blood

2.20.1 Isolation and Lysis of White Blood Cells

NH_4Cl:Tris solution was warmed to 37° C. 5 volumes of this solution was added to whole blood in 50 ml polypropylene centrifuge tubes and incubated at 37° C for 5 minutes. Samples were centrifuged for 10 minutes at 2000 rpm. The supernatant was aspirated leaving 4-5 ml above the pellet. The pellet was resuspended in 10 ml saline, centrifuged at 2000 rpm for 10 minutes, the supernatant aspirated and the saline wash repeated. The cells were resuspended in 2 ml high TE buffer and immediately lysed by the injection of 2 ml of lysis mixture using a 5 ml syringe with a 16-18 gauge needle. The injection of the lysis mixture sufficiently mixes the suspension to cause complete, instantaneous lysis. Lysate was stored at 4° C.

NH_4Cl:Tris solution
0.1395M NH_4Cl
0.017M Tris, pH7.65

Saline
0.85% NaCl

High TE Buffer
100mM Tris, pH8.0
40mM EDTA, pH8.0

Lysis Mixture
100mM Tris, pH8.0
40mM EDTA, pH8.0
0.2% SDS

2.20.2 DNA Extraction

An equal volume of TE-equilibrated phenol was added to the lysate and gently mixed by inversion to a milky white emulsion. The sample was placed on a rotator and gently mixed for

10 minutes, followed by centrifugation at 2000 rpm for 4 minutes. The upper aqueous phase was removed with a large bore pipet made by cutting the tip off of a Pasteur pipet and fire polishing it. The phenol extraction was repeated until no interphase remained. An equal volume of 24:1 chloroform:isoamyl alcohol was added, the sample placed on a rotator for 10 minutes and centrifuged for 10 minutes at 2000 rpm. The upper aqueous layer was removed to a 15 ml tube.

2.20.3 DNA Precipitation

The volume of the aqueous layer was determined and a 1/10 volume of 4 M ammonium acetate was added. An equal volume of isopropanol was added, the tube swirled to entangle the DNA into a small ball and the DNA removed from the solution using a curved end Pasteur pipet. The DNA was washed using a stream of 70% ethanol, allowed to briefly air dry and dissolved in 1 ml low TE buffer. The DNA was allowed to dissolve overnight on the rotator at room temperature. The concentration of DNA was estimated by running 1 µl in 5 µl stop buffer into a 0.8% agarose gel.

Low TE Buffer
10 mM Tris, pH8.0
1 mM EDTA, pH8.0

2.21 Epstein Barr Transformed Cell Line

Epstein Barr transformed lymphoblastoid cell lines, established from patients G.S. (MGV-280), T.P. (GM13540, Coriell Institute), and the Dhooge et al. (1994) patient (MGV-292) were maintained in continuous suspension culture. G.S. carries a de novo inversion duplication (8)(p12→p23.1) (Dill et al. 1987), T.P carries a de novo inversion duplication (8)(p11.2→p23) (table 1), and the Dhooge et al. (1994) patient carries a maternally inherited distal 8p duplication

of undetermined band origin and orientation (direct or inverted). Cells were grown in RPMI 1640 medium (StemCell Technologies) supplemented with 15% fetal bovine serum (Gibco/BRL). Cultures were split bi-weekly to maintain optimum growth.

2.22 Metaphase Chromosome Preparation

2.22.1 Pre-harvest Treatments

Epstein Barr transformed lymphoblastoid cell line cultures were harvested 2 days after sub-culturing. Chromosomes were prepared according to standard protocols (Verma and Babu, 1989) with some modifications. 15-17 hours prior to harvest 1/100 volume methotrexate (1.25 μg/ml) was added to cell culture to synchronize cell growth by arresting cells in S phase. 4 ½ to 5 hours prior to harvest, cells were rescued from arrest by removing as much medium as possible, replacing with fresh pre-warmed medium, and adding 1/100 volume bromodeoxyuridine (1 mg/ml). Thirty minutes prior to harvest, all but 10 ml of medium was removed and 0.5 ml colcemid (10 μg/ml) was added.

2.22.2 Chromosome Harvest

Cell cultures were transferred to 15ml centrifuge tubes and centrifuged for 10 minutes at 200 g. The supernatant was removed and the pellet was thoroughly resuspended. 8 ml of prewarmed 0.75 M KCl hypotonic solution was added, the cells were carefully resuspended by swirling, and placed at 37° C for 5-10 minutes. 2-3 ml 3:1 methanol:glacial acetic acid fixative was added to each tube, dropwise, with swirling. Cells were pelleted by centrifugation at 200 g for 10 minutes, the supernatant removed and the cell pellet resuspended. The cells were washed a

further 3 times in 5 ml fixative. After the final wash the cells were resuspended in 10 ml fixative and stored at -20° C.

2.23 Slide Preparation and Storage

Fixed cells were pelleted by centrifugation for 10 minutes at 200 g, the supernatant removed and the cells resuspended in enough fixative to result in a slightly cloudy mixture. Pre-washed slides stored at 4° C in 95% ethanol were rinsed multiple times in distilled water to remove all traces of ethanol and result in a smooth sheen to the slide when removed from the distilled water. A small volume of the cell suspension was taken up with a Pasteur pipet and a single drop was dropped onto one end of the slide held at a 45° angle. The solution was allowed to move down the length of the slide and then flicked to move the solution across the width of the slide. Slides were allowed to dry and were examined under phase contrast to assess quality. Slides were used immediately or were stored in 70% ethanol at -20° C (Jauch et al., 1990).

2.24 Fluorescence In Situ Hybridization (FISH)

One use of fluorescence in situ hybridization is to allow direct visual localization of a human DNA clone on metaphase chromosomes. This is particularly useful in investigations of aberrant chromosomes in patient metaphases.

2.24.1 Nick Translation Labeling of Probes

Hybridization probes were labeled with biotin by nick translation using the Gibco/BRL nick translation kit. To 1 μg of DNA was added 5 μl Solution A4, 1 μl 1 mM biotinylated dUTP (Boehringer Mannheim), and dH_2O to 45 μl. 5 μl of Solution C (DNA Polymerase/DNaseI) was added and labeling was allowed to proceed for 1 ½ hours at 14° C. 5 μl of Solution D (300mM

Na_2EDTA, pH8.0) and 1.25 µl 5% (w/v) SDS was added to stop the reaction. Probes were mixed with 67.84 µg human placental DNA (Sigma) and dH_2O to 100 µl. 1/10 volume of 3 M ammonium acetate was added. 2 volumes of 95% ethanol were added to precipitate the DNA and the solution was placed at -70° C for 30 minutes to overnight. DNA was pelleted by centrifugation, washed two times with 70% ethanol, dried and resuspended in 8 µl 1X TE (final concentration approximately 100 ng/µl)

Solution A4
0.2mM each dATP, dCTP, dGTP
500mM Tris-HCl, pH7.8
50mM $MgCl_2$
100mM 2-mercaptoethanol

2.24.2 Pre-Hybridization Slide and Probe Preparation

Slides were pre-treated with 2X SSC at 37° C for 30 minutes followed by dehydration through a room temperature ethanol series (70%, 85%, 95%, 2 minutes each). Denaturation was carried out in 70% formamide/2X SSC, pH 7.0 at 74° C for 2 minutes followed by dehydration through an ice cold ethanol series (70%, 85%, 95%, 2 minutes each). Hybridizations using cosmid 153G8 were carried out by Stanya Jurenka. 100 ng of labeled probe was mixed with 10 µl of Hybrisol VII (Oncor) and denatured at 74° C for 10 minutes. 1 µl of D8Z1 α-satellite centromere probe (Oncor) was added to the hot solution. This allows some denaturation of the repetitive centromere DNA but allows clearer visualization of both probes than if the Oncor probe is denatured completely. For biotin labeled BAC 223B23, 100 ng of probe was mixed with 5 µl of Hybrisol VII (Oncor) and denatured at 74°C for 10 minutes. The denatured probe was placed at 37° C for 1 ½ hours to allow preannealing of highly repetitive sequences. Just prior to hybridization, 5 µl of Hybrisol VII (Oncor) was heated at 74° C for 5 minutes, 1 µl of

D8Z1 α-satellite centromere probe (Oncor) was added to the hot solution, mixed, the solution was allowed to cool slightly and was mixed with the BAC 223B23 probe.

2.24.3 Hybridization

The Hybrisol VII (Oncor) diluted probe was pipeted onto the slide and a small coverslip was placed over it. The coverslip was sealed with Elmers Rubber Cement and placed in a humid chamber at 37° C overnight.

2.24.4 Post-Hybridization Washes

The coverslip was removed and the slide placed in 50% formamide/2X SSC, pH7.0 for 15 minutes at 43° C followed by 8 minutes in 2X SSC at 37° C. The slide was washed three times in 1X PBD for two minutes.

1X PBD
32.57 g $NaH_2PO_4 \cdot dH_2O$
17.3 g Na_2HPO_4 (anhydrous)
9 g NaOH
2 ml Triton X
dH_2O to 2L

2.24.5 Detection, Microscopy, and Photography

Hybridized probes were visualized using FITC-avidin (FA) which has a strong affinity for biotin and amplified using cycles of biotinylated anti-avidin D (BAAD) followed by FA. 100 µl of antibody diluted in PMN (5 µg/ml) was pipeted onto wet slides, covered with a plastic coverslip and placed at 37°C for 5 minutes. Excess antibody was removed by three - 2 minute washes in 1X PBD. The procedure was repeated, in the order: FA→BAAD→FA. The DNA was counterstained using propidium iodide (PI) mixed with antifade (0.6 µg/ml) and visualized

using a Zeiss fluorescence microscope, equipped with a propidium iodide/FITC specific filter set (excitation BP450-490, dichroic FT510, barrier LP520). Photographs were taken with 1600 ASA colour film (Fuji) and processed.

1X PMN
50 ml 1X PBD
2.5 g skim milk powder
place at 37° C overnight, pellet sediment, filter sterilize supernatant
plus 50 µl Na azide (0.2% w/v), store at 4° C

Chapter 3: Results

3.1 Polymorphic STS Genotyping

STS genotyping at D8S503 (figure 3, table 2) was carried out on 9 inv dup(8p) patients previously reported by Floridia *et al.* (1996) to investigate whether this marker is located in the single copy region of the inv dup(8p) chromosome. Marker D8S503 is located on the sex-average genetic map (Dib *et al.*, 1996, Gyapay *et al.*, 1994) proximal to D8S349 (figure 3) which is deleted from the inv dup(8p) chromosome (Floridia *et al.*, 1996). If this marker is included in the single copy region then D8S503 is the most informative distal marker in this region.

Nine inv dup(8p) patients reported by Floridia *et al.* (1996) as, 10 (CM), 13 (TS), 9 (CI), 14 (MM), 2 (PJ), 5 (SA), 15 (TG), 1 (GL), and 7 (AM), and renamed patients 1 (CM), 2 (TS), 3 (CI), 4 (MM), 5 (PJ), 6 (SA), 7 (TG), 8 (GL), and 9 (AM) for this report, were genotyped at marker D8S503 (primer sequences and annealing temperature listed in table 2). All nine patients were previously shown to have inherited the aberrant chromosome from their mother (Floridia *et al.*, 1996). Patients 2, 3, 6, 7, 8, and 9, were informative for the inheritance of a single D8S503 allele from each parent, confirming that D8S503 is present on the aberrant chromosome and therefore must be located in the single copy region. Patients 1, 4, and 5, were informative at D8S503 for the inheritance of a single paternal allele and at least one maternal allele, confirming that D8S503 is present on the aberrant chromosome. Raw genotyping data for D8S503 is presented in figure 8 and summarized in table 3. Genotyping results reported by Floridia *et al.*(1996) for D8S201, D8S349, D8S252, D8S265, D8S552 and D8S135, for these 9 patients are presented in table 3 to allow comparisons to be made to patients 10-14 reported in section 3.12.

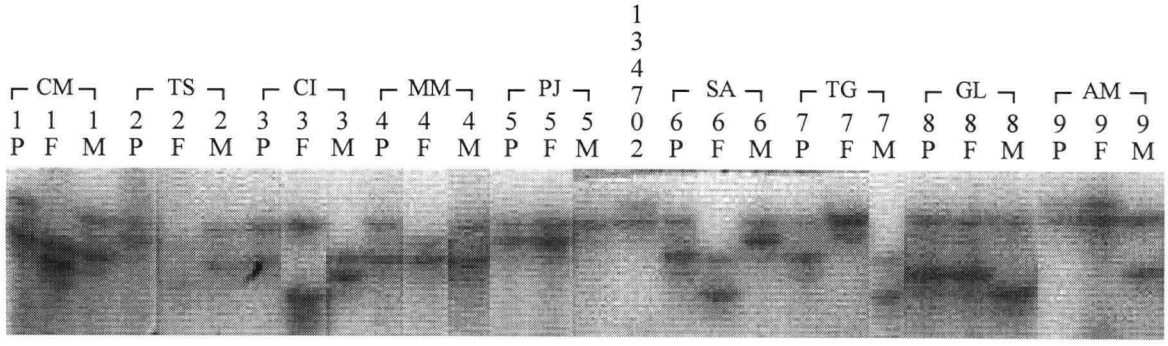


Figure 8: Genotyping of 9 families with STS polymorphic marker D8S503.
Numbers represent patient number, initials are those of the patient, P=proband, F=father, M=mother. 134702 is a *CEPH* family member included as a positive control.
The autoradiograph is a composite of two different exposure times.

Table 3: Summary of polymorphic STS genotyping of inv dup(8p) patients 1-9.

a. STS genotyping reported in this thesis. b. STS genotyping reported in Floridia et al., 1996 (patient numbers are those used in this thesis). c. Patient inversion duplication chromosome information. STS markers are listed from telomere to centromere. D8S503 is placed by YAC content, between D8S252 and D8S265. Within each triad, allele sizes are reported as A-D, A being smallest, D being largest. Genotypes at each marker are presented in the order patient, father, mother.

a.

	1. CM	2. TS	3. CI	4. MM	5. PJ	6. SA	7. TG	8. GL	9. AM
D8S503	BC AB BC	CD AC BD	CD AD BC	AC AB AC	AB AB B	BD AB CD	BC C AB	BC BC AC	BC BC AB

b.

	1. CM	2. TS	3. CI	4. MM	5. PJ	6. SA	7. TG	8. GL	9. AM
D8S201	C C AB	B B AC	B B AC	C AC AB	C AC B	A AB CD	A AD BC	B B AC	B AB C
D8S349	A AB C	A AC BC	A AB A	A AB BC	C C AB	C CD AB	A AC B	A AD BC	C BC AC
D8S252	B B AB	A A A	B B AB	A A A	A A A	A A A	B AB AB	AB AB AB	A A A
D8S265	AB AB B	AB B AC	AB AB B	A AB A	A AC AB	B B AB	AC AC AB	B B AB	B AB AB
D8S552	ABC AB AC	ABD AC BD	ABC AB AC	A AB AB	BC AB C	BC AC B	ABC BC AB	BCD AC BD	BC AC BC
D8S135	BD BC AD	AB BC AC	BD BC AD	AC BC A	AB AB AB	BC AC BC	AB AB AB	AB B A	B AB AB

c.

Patient	Duplicated Region	Distal Deletion	origin
1. CM	8p11.2→8p23.1	8p23.1→8ter	maternal
2. TS	8p11.2→8p23.1	8p23.1→8ter	maternal
3. CI	8p11.2→8p23.1	8p23.1→8ter	maternal
4. MM	8p21.1→8p23.1	8p23.1→8ter	maternal
5. PJ	8cen→8p23.1	8p23.1→8ter	maternal
6. SA	8cen→8p23.1	8p23.1→8ter	maternal
7. TG	8p21→8p23.1	8p23.1→8ter	maternal
8. GL	8cen→8p23.1	8p23.1→8ter	maternal
9. AM	8p11.2→8p23.1	8p23.1→8ter	maternal

3.2 Refinement of the Map of the Region Predicted to Contain the Distal Element

For inv dup(8p) patients with a cytogenetic center of symmetry in band 8p23.1, STS genotyping confirms that in 6/6 informative patients, D8S503 is located in the region of single copy at the cytogenetic center of symmetry (table 3). Genotyping also confirms the deletion of D8S349 from the inv dup(8p) chromosome in 11/11 informative patients (Floridia *et al.*, 1996). The location of the distal repetitive element predicted by the hypothesis of a mechanism of formation mediated by inverted repeats, can then be refined to the region between marker D8S349 and D8S503. Currently this region is represented by a gap in the database of YAC singly and doubly-linked STS contigs maintained by the MIT/Whitehead Institute for Genome Research (http://www.genome.wi.mit.edu). This gap is located between contigs WC8·0 (distal) and WC8·1 (proximal). In order to investigate this region at the molecular level, this contig gap must be spanned. Mapping efforts were focused on this goal.

The MIT/Whitehead Institute data base was used to identify YACs for additional STS marker testing. The information in this database is collected by a 'pooled clone' PCR approach. Consequently, positive data must be confirmed and negative data is not reported. Thirteen YACs from contigs WC8·0 and WC8·1, comprising the proximal and distal ends respectively, were chosen for analysis (YACs 4-16, table 4). With the exception of 856d8, 843e1, 889b7, and, 966b5, the YACs from this region were known to be chimeric, containing segments of DNA from more than one region of the genome. Chimeric YACs are useful for STS content contig assembly but are not useful for chromosome walking. STS content analysis, by PCR, of YACs 4-6 chosen from WC8·0, was carried out for markers D8S439, D8S277, D8S349, D8S349 (the same $(CA)_n$ as D8S1819), and D8S1935. STS content analysis by PCR, of YACs 7-16 chosen

from WC8·1, was carried out for markers D8S252, D8S503, D8S1825, D8S574, D8S516 and D8S542. All primer sequences and annealing temperatures are listed in table 2. A comparison of the MIT/Whitehead Institute STS content mapping, and results obtained in the chosen subset of YACs tested, is presented in table 5 and discussed below. STS content data collected for this thesis is summarized in figure 9. An example of the raw PCR data is presented in figure 10.

3.2.1 YAC STS Content Not Reported by MIT/Whitehead Institute

For the region of WC8·0 studied all MIT/Whitehead positive results for the STS markers D8S277, D8S1819 and D8S1935 were confirmed. YAC 764c7 was extended by the presence of markers D8S277 and D8S439. YACs 967c11 and 856d8 were found to contain marker D8S439.

For the region of WC8·1 studied, almost all MIT/Whitehead positive results for the STS markers tested were confirmed (see section 3.2.2). YACs 966b5 and 889b7 were found to contain D8S1825. YACs 742d12 and 843e1 were found to contain D8S542. YACs 845a9, 849h4, and 785d6 were found to contain D8S503. YAC 845a9 was found to contain D8S516. Markers D8S252, D8S574, and D8S351, not analysed by the MIT/Whitehead Institute, were also tested for their presence in these YACs. Marker D8S252 is present in YACs 785d6, 742d12, and 889b7. Marker D8S574 is present in YACs 849h4, 871a8, 742d12, and 966b5. Marker D8S351 is present in YACs 785d6, 966b5, and 843e1.

3.2.2 Negative YAC STS content Not Predicted by MIT/Whitehead Institute

With the exception of 966b5, which did not contain D8S516 when tested in our lab, all results were in agreement with those predicted by MIT/Whitehead. This negative result could reflect a false positive in the MIT/Whitehead data as the frequency of false positives in these

Table 4: YAC information.

The subset of YACs investigated in this thesis have been placed on chromosome 8 by STS content (MIT/Whitehead Institute). n.k. = size is not known. "-"= YAC is not known to be chimeric.

YAC	SIZE(kb)	chimeric with chromosome:
1. 810f8	1180	9
2. 741h4	n.k.	14
3. 787c11	1740	-
4. 967c11	1110	17
5. 856d8	550	-
6. 764c7	1370	17
7. 918b5	1320	2
8. 849h4	1380	2
9. 871a8	500	11
10. 776f4	1740	15
11. 785d6	1320	1,10
12. 742d12	n.k.	2
13. 966b5	1680	-
14. 889b7	n.k.	-
15. 845a9	600	2
16. 843e 1	580	-
17. 915h4	740	-
18. 773g4	1570	17
19. 799b1	480	-

Table 5: Summary of YAC STS content data.

STS markers are listed from left to right in order from telomere to centromere. STS marker order is based on YAC content analysis and placement on the Genethon genetic map. "+" = Presence of an STS marker in a YAC clone as reported by the MIT/Whitehead Institute. "*" = Presence of an STS marker in a YAC clone confirmed in this thesis. "-" = absence of an STS marker in a YAC clone reported in this thesis

	D8S439	D8S277	D8S349	D8S1935		D8S252	D8S503	D8S1825	D8S574	D8S351	D8S516	D8S542
YAC												
4. 967c11	*	+ *	+ *	+ *								
5. 856d8	*	+ *	+ *	-								
6. 764c7	*	*	+ *	-								
7. 918b5								+ *				
8. 849h4							*	+ *	*			
9. 871a8								+ *	*			
10. 776f4							+ *	+ *				
11. 785d6						*	*	-	-	*	+ *	
12. 742d12						*	+ *	*	*	-	+ *	*
13. 966b5							+ *	*	*	*	+ -	
14. 889b7						*	+ *		-	-	-	-
15. 845a9						-	*		-	-	*	-
16. 843e1						-	-		-	*	+ *	*

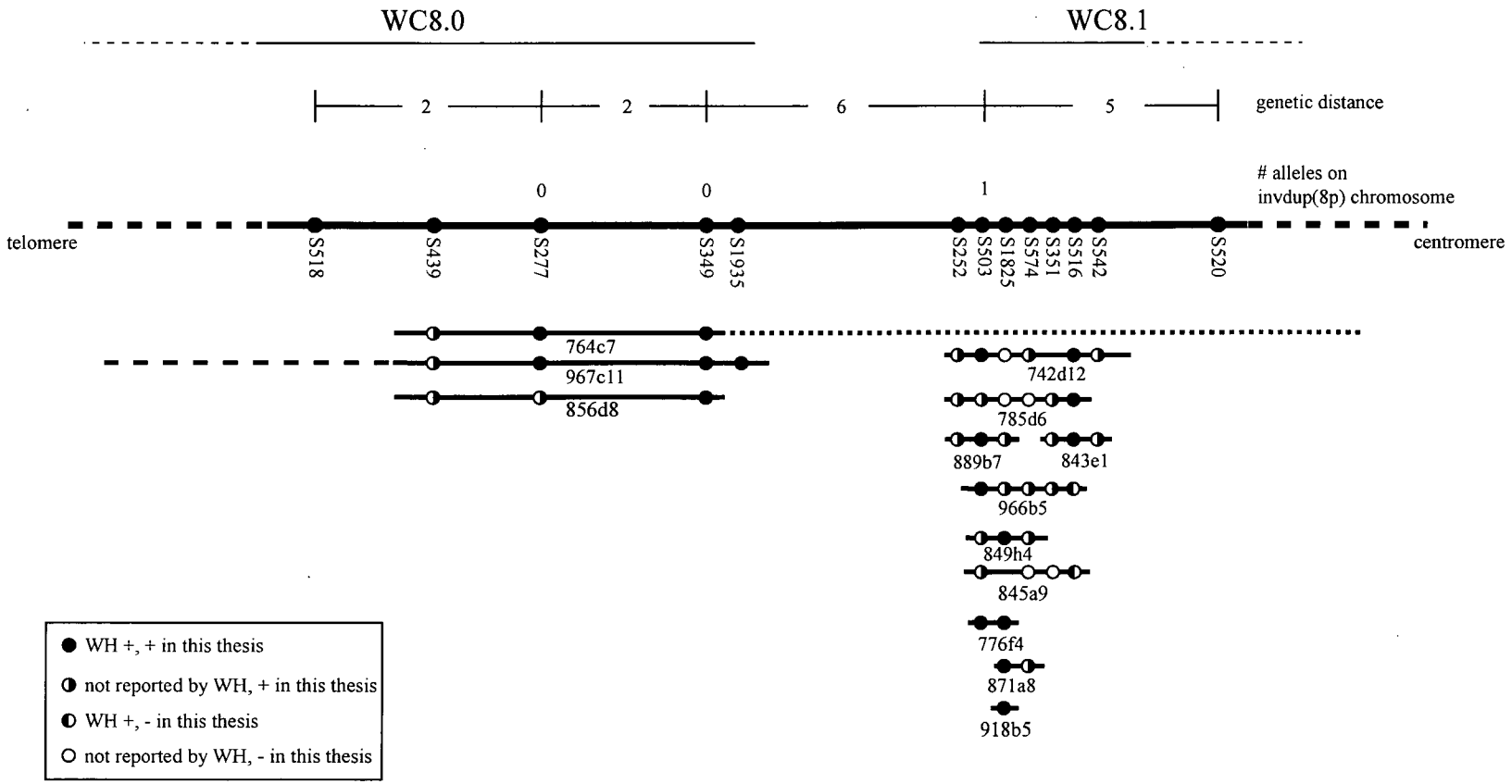
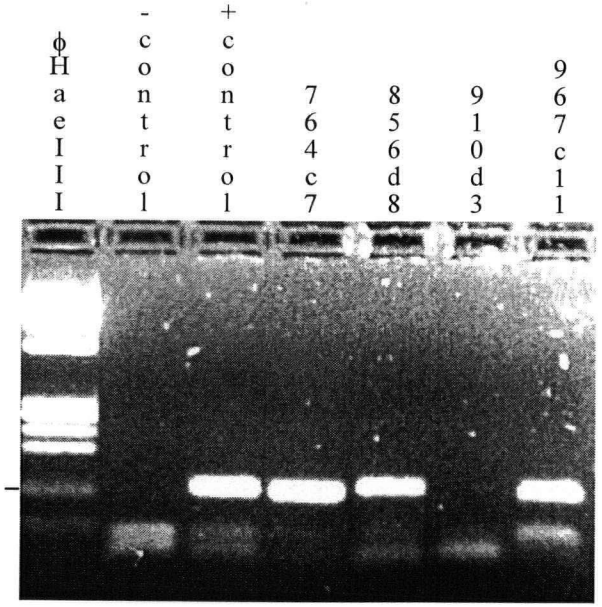


Figure 9: STS content mapping in YACs. Comparison of MIT/Whitehead Institute (WH) STS content data for YACs selected from contigs WC8.0 and WC8.1 to STS content data reported in this thesis. Modifications, from those reported by the MIT/Whitehead Institute, have been made to STS marker order as a result of testing with additional STS markers. Genetic distance is reported in centimorgans.

a. D8S439 - size: 115-133



b. D8S542 - size: 229-249

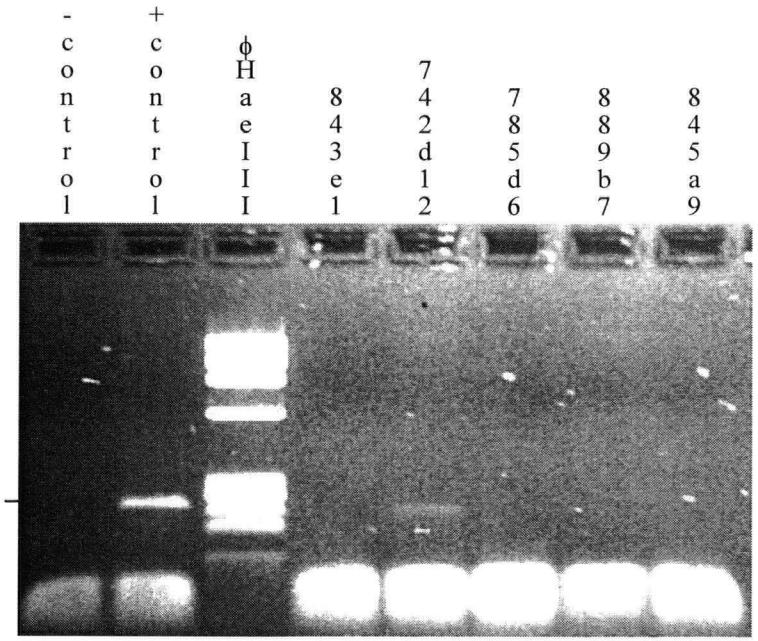


Figure 10: Examples of STS content analysis by PCR. All clones are YACs. (+) control=chromosome 8 hybrid cell line MGV-269 (-) control= milliQ filtered distilled water. ϕHaeIII is included as a size marker. PCR amplification was carried out for: a. STS polymorphic marker D8S439, b. STS polymorphic marker D8S542.

data is known to be 6% (Hudson *et al.*, 1995). However, given that large YACs are prone to deletion, the possibility also exists that a deletion has occurred in our clone that wasn't present in the CEPH mega-YAC pools.

STS content analysis did not lead to closure of the gap between WC8.0 and WC8.1. A variety of approaches were undertaken to isolate clones spanning this region but, were unsuccessful (data not shown). Efforts were turned to analysis of a candidate repetitive element.

3.3 Isolation of an 8p Repetitive Element

A novel repetitive element, termed a megasatellite, was isolated from 4p15 and localized to 4p15 and 8p23 by Southern blot analysis of single chromosome hybrids and by FISH analysis (Gondo *et al.*, 1996, Kogi *et al.*, 1997). The 4p element is a 4.7 kb *Eco*RI fragment, tandemly repeated 12-90 times in the Japanese population (Gondo *et al.*, 1996, Kogi *et al.*, 1997). The localization of this element to distal 8p, and the low copy number in the genome, made it a candidate for the proposed repetitive element involved in the formation of inv dup(8p) chromosomes. Chromosome 8 megasatellite clones were isolated and their organization on the chromosome investigated.

The 4753 bp sequence of Gondo et al., 1996 (Genbank accession #: D38378) was used as a query sequence in a BLASTsts search of the Genbank STS database (http://www.ncbi.nlm.nih.gov). STS markers rsAVA13 (accession #U57857), STS4-310 (accession #L00791), SHGC4-672 (accession #G01777), SHGC4-1030 (accession #G01800), SHGC4-1135 (accession #G01806), SHGC4-1596 (accession #G01912), SHGC4-1739 (accession #G02001), SHGC4-1436 (accession #G01841), and SHGC4-1656 (accession #G05153) were recovered. Homology of these markers to the chromosome 4 megasatellite

sequence are shown in figure 11. rsAVA13 was derived from a chromosome 8 telomere half YAC, the others, from chromosome 4 random clones (see table 2 for references).

3.3.1 Isolation of Megasatellite Containing Cosmids

PCR products, amplified from the STSs spanning the megasatellite, were pooled and used as hybridization probes to the LA08NC01 cosmid library gridded on high density filters (See table 2 for STS markers, primer sequences and annealing temperatures. Primer pairs are available from Research Genetics, Inc.). Initially, only a single cosmid, 153G8, was isolated.

3.3.2 Characterization of Cosmid 153G8

Megasatellite STS PCR products were hybridized to *Eco*RI digested cosmid 153G8. Five *Eco*RI fragments; 4.7 kb, 3.35 kb, 2.4 kb, 2.3 kb, and 2.1 kb contain sequences which cross-hybridize with the megasatellite sequence. Fragments are identified by cosmid name (in this case, 153G8), enzyme used to create the fragment (in this case, *Eco*RI (E)), and the fragment size in kb. Therefore, these fragments would be identified as 153G8E4.7, 153G8E3.35, 153G8E2.4, and so on. STS PCR products were individually hybridized to *Eco*RI digests of 153G8: 153G8E4.7 hybridized with all of the products, 153G8E3.35 hybridized only with the 3′ products, 153G8E2.1 hybridized only with the 5′ products, and 153G8E2.3/153G8E2.4 (not resolvable in this figure) hybridized with all of the products (figure 12). The 4p megasatellite sequence was analyzed for motifs, that with a single base pair change, would create an *Eco*RI site. A mutation of G→T at 2332bp would create an *Eco*RI recognition site (GAATTC) resulting in two fragments, one of 2.3 kb and one of 2.4 kb. This would explain the apparent hybridization of all PCR products to 153G8E2.3/153G8E2.4.

The *Eco*RI blots were hybridized with γ-end labeled T7 and T3 primers (see table 2 for sequences). The T3 end of 153G8 is contained within the 4.9 kb fragment and the T7 end within the 0.5 kb fragment. These fragments do not correspond to any of the fragments cross-hybridizing with the megasatellite sequences. Therefore, the megasatellite sequences must lie internally in cosmid 153G8. As well, there are at most, 3 complete copies of the 4.7 kb *Eco*RI unit in this region of the genome. This is significantly different from chromosome 4, where the megasatellite is present in 12-90 copies.

Cosmid 153G8 was digested with *Bam*HI and *Eco*RI, blotted, and hybridized with the megasatellite STS PCR products. At the resolution obtained by digestion with *Bam*HI, all *Eco*RI fragments containing the megasatellite, with the exception of 153G8E3.35, cross hybridized with the megasatellite PCR products. The 153G8E3.35 fragment is cleaved, by digestion with *Bam*HI, into 153G8BE1.1 and 153G8BE2.2, and only 153G8BE2.2 hybridized with the megasatellite PCR products (data not shown).

A restriction enzyme map of 153G8 was assembled using the enzymes, *Eco*RI, *Bam*HI, *Bgl*II, *Bss*HII, and *Kpn*I. This map is shown in figure 13 and shows the location of the megasatellite sequences relative to the ends of 153G8. Throughout this thesis, references to megasatellite or flanking sequences with regard to 5′ or 3′ are made relative to this restriction map.

3.3.3 Characterization of Cosmid 39A7

A second screen of the cosmid library, using the pooled PCR products, isolated cosmid 39A7. Digestion of 39A7 and hybridization with the pooled megasatellite PCR products identified 5 *Eco*RI fragments identical in size to the fragments for 153G8 with the exception that the 3.35 kb fragment of 153G8 is 3.8 kb in 39A7 (figure 14). Digestion of 39A7E3.8 with *Bam*HI results

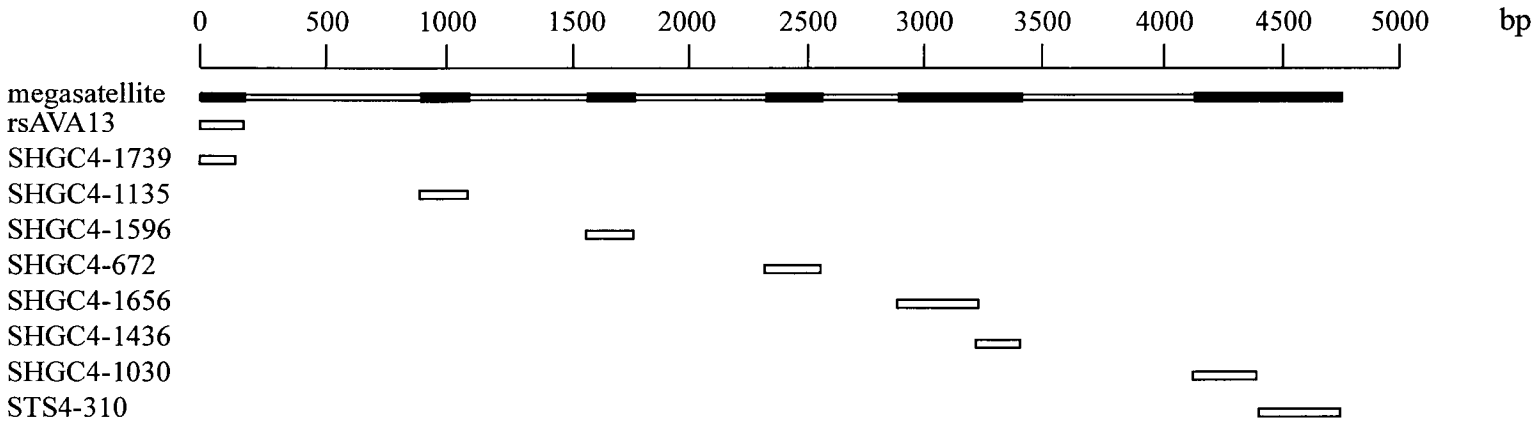


Figure 11: Homology of STS markers to 4p megasatellite sequence. The 4p megasatellite is 4746 bp in length. Blocks of homology are shown as black boxes, the homologous STS marker is shown as an open box below the region of homology. Marker rsAVA13 was derived from a human chromosome 8 clone. The remaining STSs were derived from human chromosome 4 clones.

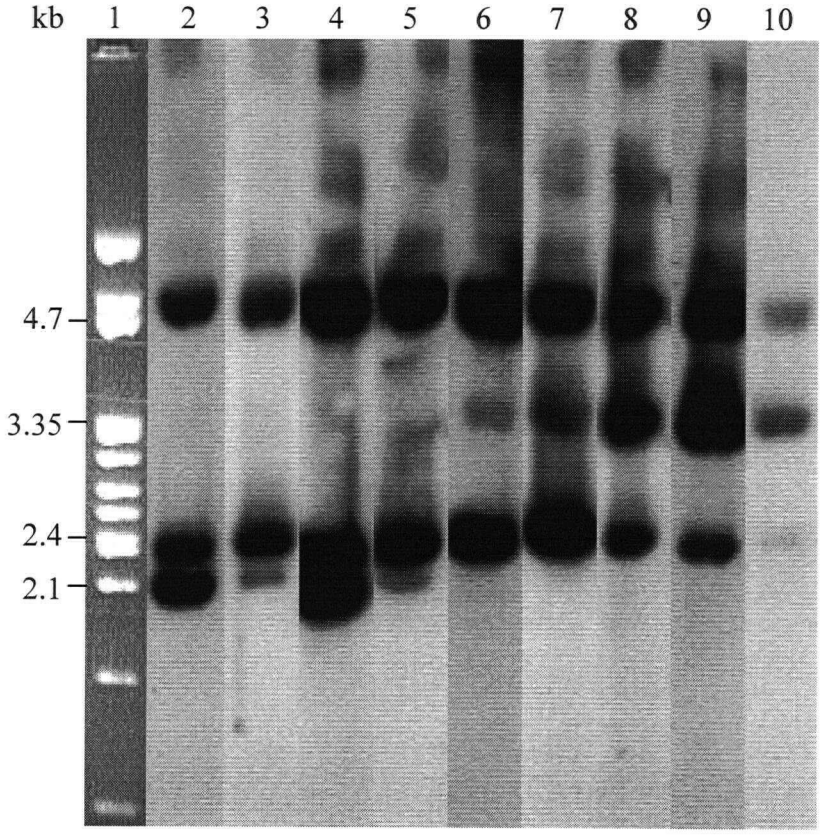


Figure 12: Hybridization of megasatellite PCR products to cosmid 153G8.
lane 1=*Eco*RI digested 153G8 banding pattern, lanes 2-10 hybridization of *Eco*RI digested 153G8 with PCR products: 2. rsAVA13, 3. SHGC4-1739, 4. SHGC4-1135, 5. SHGC4-1596, 6. SHGC4-672, 7. SHGC4-1656, 8. SHGC4-1436, 9. SHGC4-1030, 10. SHGC4-310.

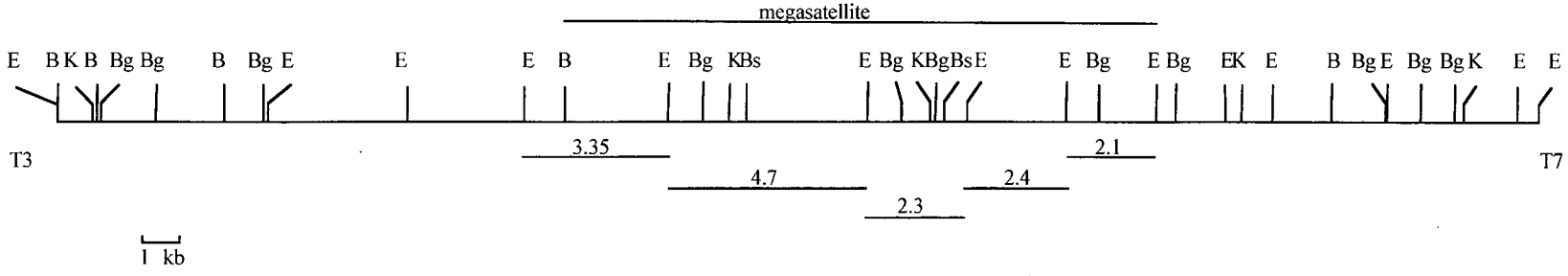


Figure 13: Restriction map of cosmid 153G8. Cosmid 153G8 contains an internal megasatellite. E=*Eco*RI, B= HI, Bg=*Bgl*II, Bs=*Bss*HII, K=*Kpn*I

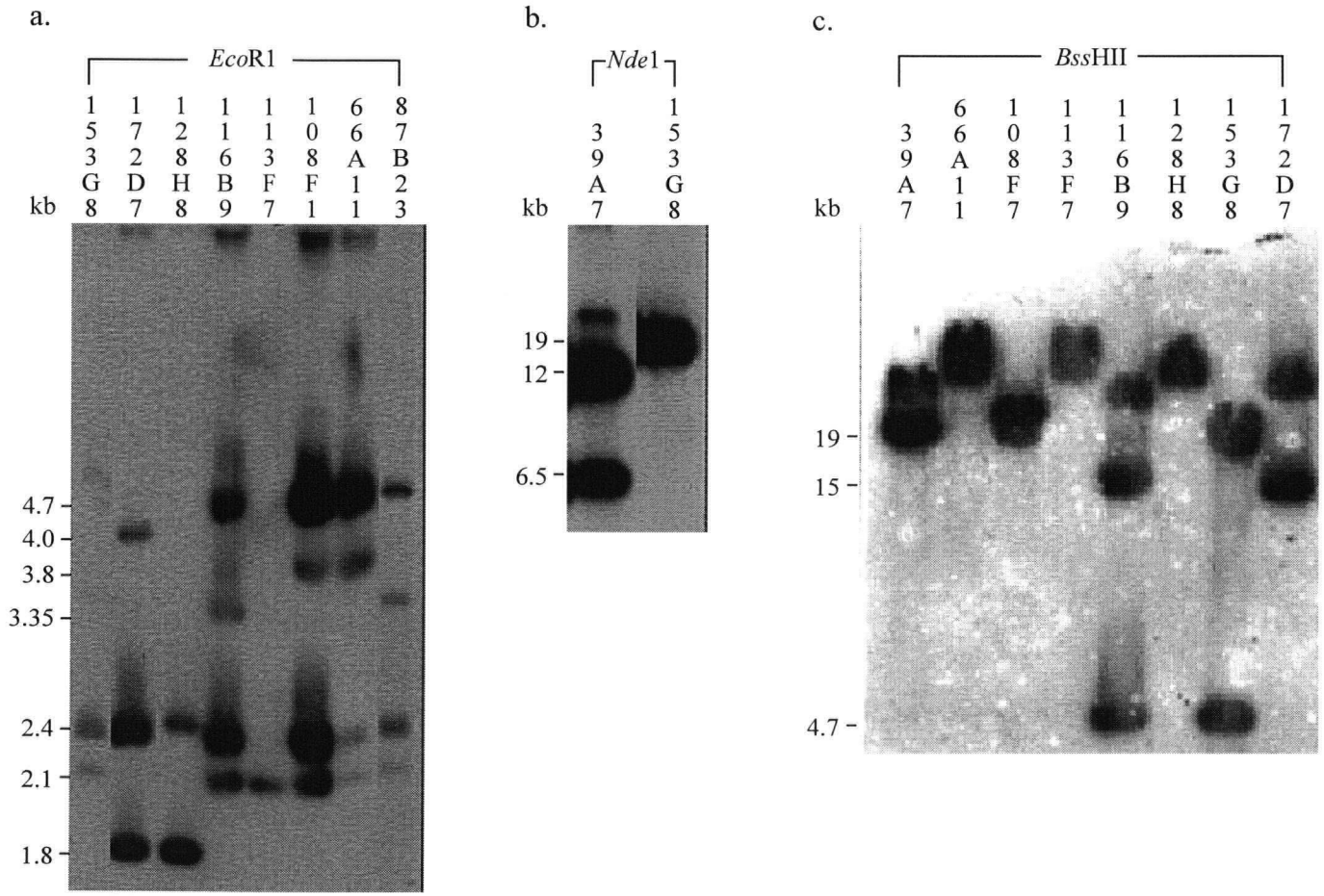


Figure 14: Characterization of megasatellite containing cosmid and BAC clones. With the exception of BAC 87B23, all clones are cosmids. Hybridization of megasatellite DNA to clones: a. digested with *Eco*RI, b. digested with *Nde*I, 3. digested with *Bss*HII. Extraneous lanes have been deleted from the autoradiographs.

in cleavage into 39A7BE1.6 and 39A7BE2.2. The megasatellite PCR products hybridize to 39A7BE2.2 but not 39A7BE1.6.

The T7 primer hybridizes to the 3.5 kb *Eco*RI fragment and the T3 primer to the 0.8 kb *Eco*RI fragment, indicating that the megasatellite must lie internally within this cosmid (data not shown). Therefore, this cosmid contains approximately three complete copies of the 4.7 kb unit. The cosmid 39A7 megasatellite can be differentiated from the cosmid 153G8 megasatellite by a difference in a single *Eco*RI fragment, 3.8 kb and 3.35 kb respectively.

3.3.4 Screening Cosmid Library with 4.7 kb Megasatellite Sequence

A final screen of the cosmid library, using the 4.7 kb megasatellite fragment isolated from both 153G8 and 39A7 rather than the pooled PCR products, led to the isolation of 6 additional cosmids; 116B9, 113F7, 66A11, 108F1, 128H8, 172D7. In order to make comparisons to 153G8 and 39A7, these cosmids were digested with *Eco*RI and hybridized with the 4.7 kb megasatellite fragment. Cosmid 116B9 has a 3.35 kb fragment, 66A11 and 108F1 has a 3.8 kb fragment, 113F7 has too little of the megasatellite sequences to determine this fragment size, and 128H8 and 172D7 have a novel 1.8 kb fragment (figure 14). The LA08NC01 cosmid library was created from a single chromosome 8, therefore, differences in restriction fragment pattern must reflect the presence of at least 3 types of the megasatellite on chromosome 8, rather than polymorphism.

3.3.5 Restriction Fragments Characteristic of Type I, II, and III Megasatellite

In order to characterize each megasatellite type, comparative restriction enzyme digestions were carried out on the cosmids. Restriction enzymes were chosen based on analysis of the

sequenced 4p megasatellite. Restriction enzymes were used in both single and double digestions, the digested DNA was transferred to nylon membrane, and hybridized with the 4.7 kb megasatellite fragment (results not shown). Figure 14 shows an example of these experiments. Ultimately, characteristic restriction fragment sizes were determined for each megasatellite type (table 6). Type I megasatellite is characterized by the presence of a 3.35 kb *Eco*RI fragment, presence of a 4.7 kb *Bss*HII fragment, and of a single, 19 kb *Nde*I fragment. Type II megasatellite is characterized by the presence of a 3.8 kb *Eco*RI fragment, absence of a 4.7 kb *Bss*HII fragment, and presence of two *Nde*I fragments, one large and one 6.5 kb in size. Type III megasatellite was initially characterized only by the presence of a 1.8 kb *Eco*RI fragment, as both cosmids containing this type did not contain a complete 4.7 kb fragment, and therefore, appeared to terminate within the megasatellite. However, later experiments with YAC clones indicate that both the 1.8 kb and 4.0 kb fragment are characteristic of the type III megasatellite. Cosmids 153G8 and 116B9 contain type I, 39A7, 108F1, 66A11 contain type II, 172D7 and 128H8 contain type III and, 113F7 contains type I or type II (figure 15).

3.4 Subcloning of *Eco*RI Fragments from Cosmids 153G8 and 39A7

The megasatellites found in cosmids 153G8 and 39A7 are internal. These cosmids were chosen for subcloning in order to isolate fragments flanking the megasatellites for analysis. A number of fragments from each cosmid were subcloned (table 7). The fragments were liberated from the cloning vector, and labeled for use as individual hybridization probes to *Eco*RI digests of all of the cosmids. Results of these hybridizations are summarized in table 7. The sequence flanking all three types of megasatellite cross hybridized, including the 4.9 kb and 0.5 kb end

Table 6: Restriction enzyme fragment sizes characteristic of the 8p megasatellite sequences.

The type I megasatellite sequences are distinguished by a 3.35 *Eco*RI fragment and a 4.7 kb *Bss*HII fragment. The type II megasatellite sequences are distinguished by a 3.8 kb *Eco*RI fragment and a 6.5 kb *Nde*I fragment. The type III megasatellite sequences are distinguished by a 4.0 kb and a 1.8 kb *Eco*RI fragment.

Megasatellite	***Eco*RI fragments**	***Nde* I fragments**	***Bss*HII fragments**
Type I	3.35 kb	1 large	2 large, 4.7 kb
Type II	3.8 kb	1 large, 6.5 kb	2 large
Type III	4.0 kb, 1.8 kb	1 large	2 large

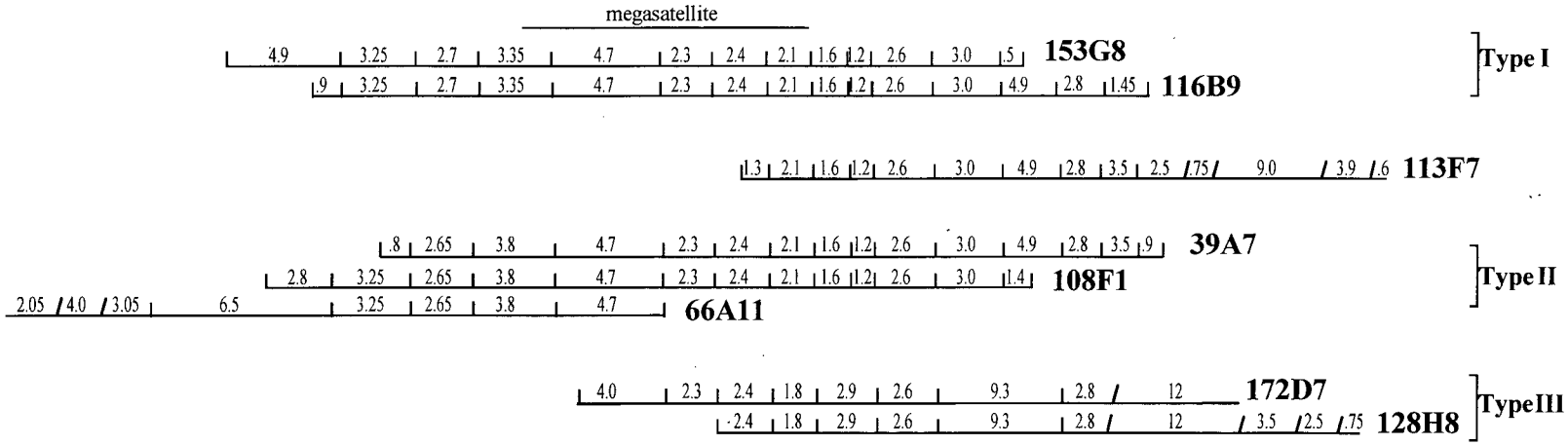


Figure 15: Homology of megasatellite containing cosmids. Type I, II and III megasatellite cosmids are from different regions of the genome. Homology of different types is represented by overlaps but is not meant to imply true overlap in the genome. Fragment sizes are in kb and result from digestion with *Eco*RI. Forward slashes represent unknown fragment order. Cosmid 113F7 is of unknown type, but is type I or II based on restriction pattern.

Table 7: Hybridization of randomly cloned *Eco*RI fragments to megasatellite containing cosmids.

Fragments were randomly cloned from cosmids 153G8 and 39A7. Megasatellite containing cosmids were digested with *Eco*RI, transferred to nylon membrane, and hybridized with: a. *Eco*RI fragments of cosmid 153G8. b. *Eco*RI fragments of cosmid 39A7. All fragment sizes are in kilobases. NH = no hybridization to this cosmid.

a.

153G8 *Eco*RI hybridization fragment	**4.9**	**3.3**	**2.7**	**1.6**	**2.6**	**3**	**0.5**
Hybridizes to 153G8 *Eco*RI fragment:	4.9	3.25	2.7	1.6	2.6	3	0.5
Hybridizes to 116B9 *Eco*RI fragment:	0.9	0.9	2.7	1.6	2.6	3	4.9
Hybridizes to 113F7 *Eco*RI fragment:	NH	NH	NH	1.6	2.6	3	4.9
Hybridizes to 39A7 *Eco*RI fragment:	NH	0.8	2.65	1.6	2.6	3	4.9
Hybridizes to 108F1 *Eco*RI fragment:	2.8	3.25	2.65	1.6	2.6	3	NH
Hybridizes to 66A11 *Eco*RI fragment:	6.5	3.25	2.65	NH	NH	NH	NH
Hybridizes to 172D7 *Eco*RI fragment:	NH	NH	NH	2.9	2.6	9.3	9.3
Hybridizes to 128H8 *Eco*RI fragment:	NH	NH	NH	2.9	2.6	9.3	9.3

b.

39A7 *Eco*RI hybridization fragment:	**1.6**	**1.2**	**2.6**	**3**	**2.8**
Hybridizes to 153G8 *Eco*RI fragment:	1.6	1.2	2.6	3	NH
Hybridizes to 116B9 *Eco*RI fragment:	1.6	1.2	2.6	3	2.8
Hybridizes to 113F7 *Eco*RI fragment:	1.6	1.2	2.6	3	2.8
Hybridizes to 39A7 *Eco*RI fragment:	1.6	1.2	2.6	3	2.8
Hybridizes to 108F1 *Eco*RI fragment:	1.6	1.2	2.6	3	NH
Hybridizes to 66A11 *Eco*RI fragment:	NH	NH	NH	NH	NH
Hybridizes to 172D7 *Eco*RI fragment:	2.9	2.9	2.6	9.3	2.8
Hybridizes to 128H8 *Eco*RI fragment:	2.9	2.9	2.6	9.3	2.8

fragments of 153G8. A pictorial representation of this is shown in figure 15 where regions of cross hybridization are represented as overlapping although the cosmids are not believed to truly overlap in the genome. Cosmid inserts containing each type of megasatellite were labeled and each type was hybridized back to cosmids and a BAC containing a megasatellite (figure 16) (see below for data on the BAC). Some differences in restriction fragment size are apparent, however, cross hybridization was demonstrated across at least 40 kb. Therefore, it appears as though each class of megasatellite is embedded within a large reiterated sequence (LRS).

3.5 Sequencing of Clones 153G8E3.35 and 39A7E3.8

The megasatellites of 153G8 and 39A7 differ in their 5′ *Eco*RI restriction fragments. These differences may offer clues to the evolution of the megasatellite sequences on chromosome 8. The 5′ *Eco*RI restriction fragments, 153G8E3.35 and 39A7E3.8, are both cleaved by *Bam*HI into two fragments. In each case only one of the fragments, 153G8BE2.2 and 39A7BE2.2 respectively, hybridize with the megasatellite sequences. The non-hybridizing fragment, 153G8BE1.1 and 39A7BE1.6, may contain novel sequences, useful as hybridization probes to study these specific megasatellite types in inv dup(8p) patient.

In order to sequence in both directions from the *Bam*HI site located within each *Eco*RI fragment, deletion clones 153G8BE1.1, 153G8BE2.2, 39A7BE1.6, and 39A7BE2.2 were constructed by the procedure outlined in figure 17. The approach chosen for construction depended on in which orientation(s) the fragment had been cloned into the vector. Clones are labeled as forward or reverse to allow clones of different orientation to be distinguished from one another. Sequencing from both ends of 153G8E3.35 and 39A7E3.8 and in both directions from the internal *Bam*HI site in each fragment was carried out using an automated sequencer (ABI

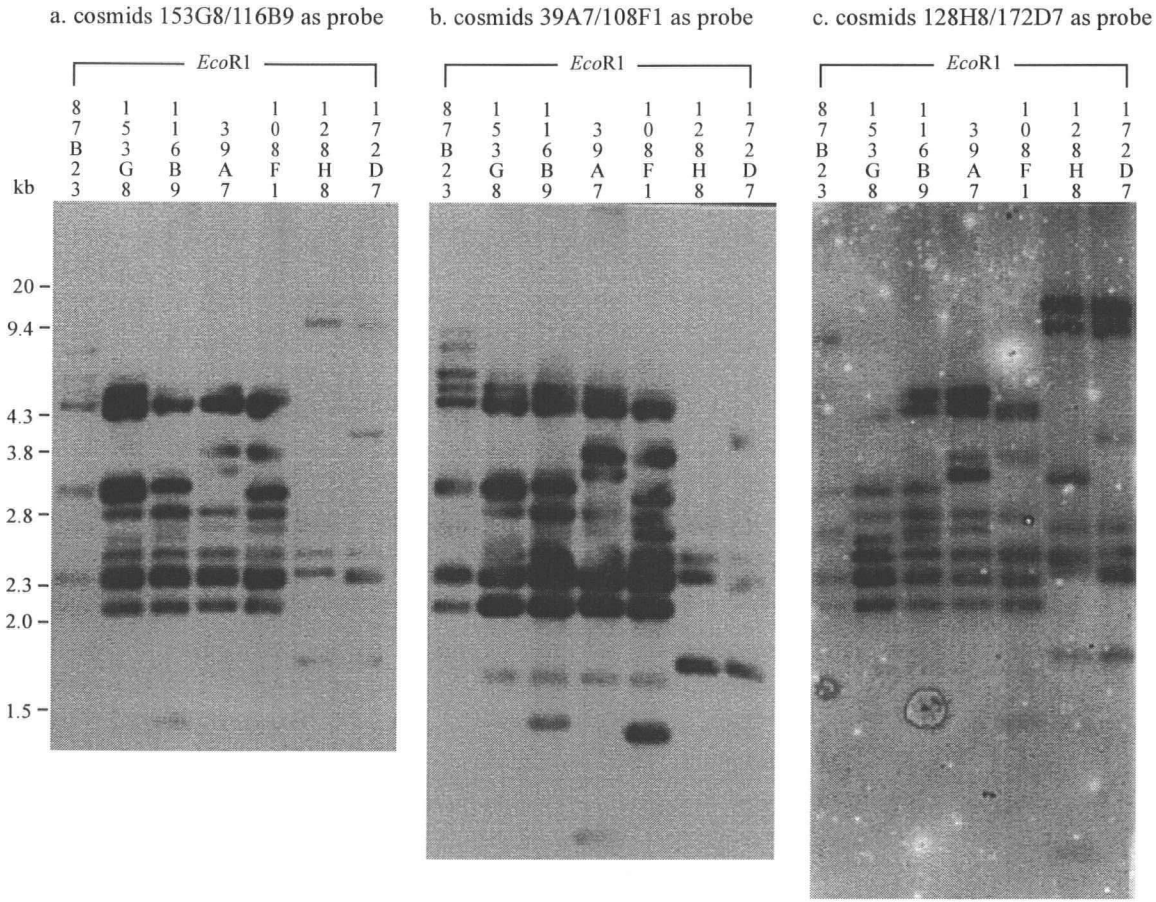


Figure 16: Cross hybridization of flanking sequences of type I, II and III megasatellites. With the exception of BAC87B23, all clones are cosmids. Clones were digested with *Eco*RI and hybridized with a. type I megsatellite cosmids, b. type II megasatellite cosmids, c. type III megasatellite cosmids.

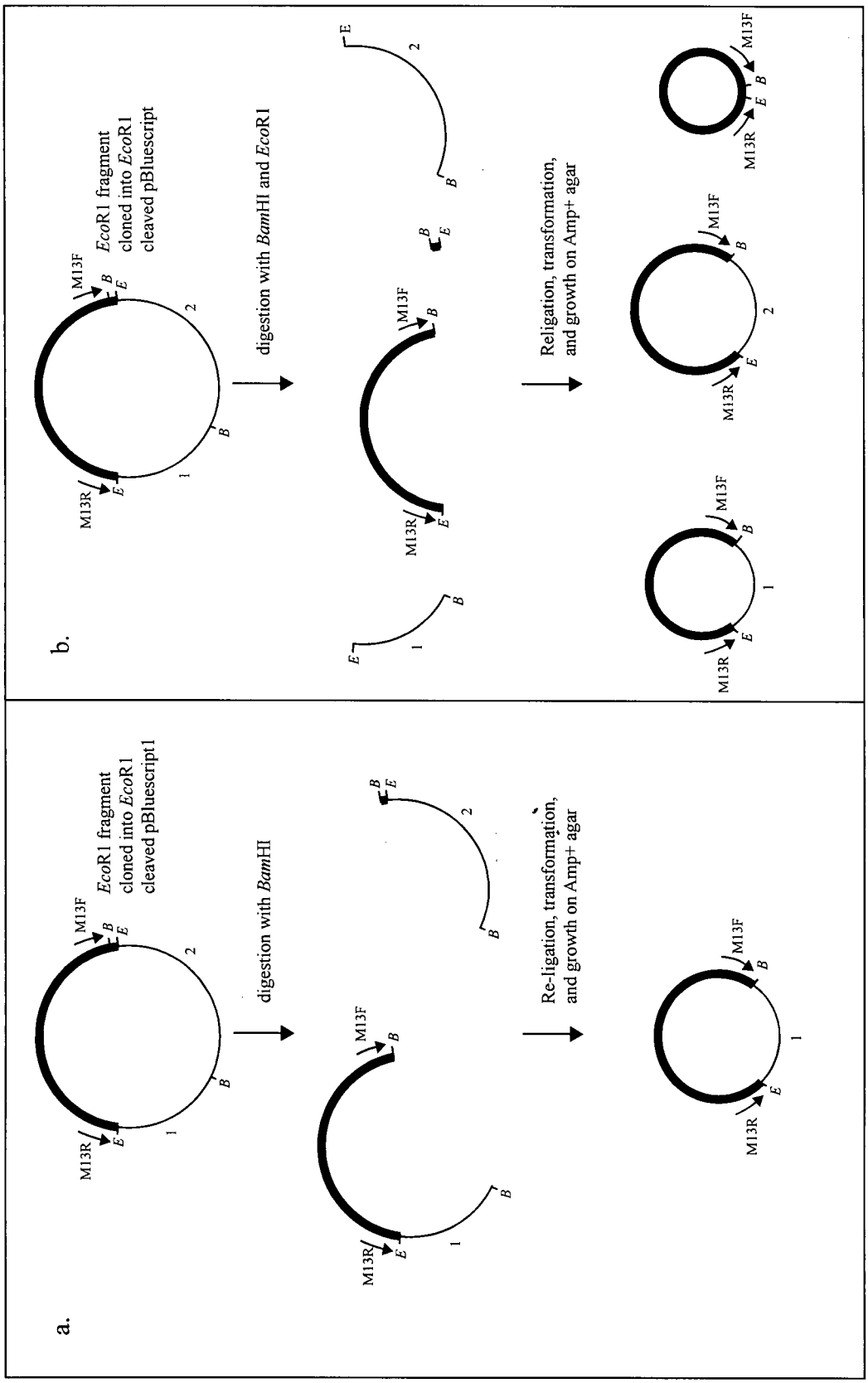


Figure 17: Construction of deletion clones from 153G8E3.35 and 39A7E3.8 F and R refer to orientation of insert within vector. a. 153G8E3.35-F and -R, 39A7E3.8-F, are digested with *Bam*HI, religated, transformed into competant *E. coli* cells and grown on selective medium. b. 39A7E3.8-F is digested with *Bam*HI and EcoR1, religated, transformed into competant *E. coli* cells and grown on selective medium. Dark line represents vector DNA, light line represents human insert. B=*Bam*HI, E=*Eco*RI.

Model 373). The regions sequenced are depicted in figure 18. Approximately 500 bp of sequence was obtained in the direction of each arrowhead. BLAST searches and Clustal W alignment were used to compare the sequences. Alignments were made between corresponding subclone sequences and the 4p megasatellite (Appendix 1). In other words, sequence from the *Eco*RI site of 153G8BE2.2 was compared to sequence from the *Eco*RI site of 39A7BE2.2 and to the 4p megasatellite, and so on. The sequences read from the *Eco*RI sites of 153G8BE2.2 and 39A7BE2.2, and from the *Bam*HI sites of these fragments, respectively, have a high degree of homology to each other and to the 4p megasatellite. 153G8BE1.1 and 39A7BE1.6 also have a high degree of homology with each other and some homology with the megasatellite, when read from both the *Bam*HI and *Eco*RI site (figure 18). This is surprising because these fragments do not hybridize with the megasatellite sequence. It is unclear whether a single 500 bp insertion/deletion, or if insertions/deletions of one or two nucleotides at multiple locations throughout the fragments, lead to the 500 bp difference in fragment size. From the number of single base pair deletions/insertions in the sequence alignments, the latter may be more likely.

3.6 Isolation of BACs Containing Megasatellite Sequences

Since cosmids containing all three types of megasatellite cross-hybridized an attempt was made to isolate larger BAC clones that might contain unique flanking sequences for each type of megasatellite. The chromosome 8 enriched BAC library was screened with the 4.7 kb megasatellite fragment. BAC 87B23 was isolated, subjected to the restriction enzyme digestions used to characterize the megasatellite, and found to contain a type I repeat. BAC 87B23 has been sized using PFGE and the insert is approximately 165 kb (data not shown). BAC 87B23 contains the STS markers D8S1619, D8S1993, and D8S1640. These 3 markers have been placed

on the chromosome 8 map, by YAC contig STS content mapping at the MIT/Whitehead Institute, to the region between D8S265 and D8S552 (figure 19). Attempts to isolate additional megasatellite containing clones from the chromosome 8 enriched BAC library were unsuccessful.

An attempt was made to isolate megasatellite containing BAC clones from the Research Genetics BAC library (Sambrook *et al.*, 1989) from which the enriched chromosome 8 BAC library was made. Hybridization of the 4.7 kb megasatellite fragment to this library would result in isolation of a preponderance of clones containing chromosome 4p15 sequences. Therefore, the Research Genetic, Inc. library was screened with the T7 and SP6 termini of BAC 87B23. An extremely large number of positive signals, estimated to be in excess of 3,000, were obtained. Only strong signals were chosen, and still over 100 positive clones were obtained. To decrease the number of clones for analysis, only those BACs contained in the 8p enriched BAC library were chosen (data listed in Appendix 2). None of these BACs contained a megasatellite.

In order to address the question of why the megasatellite may be underrepresented in the chromosome 8 BAC library, an estimate of the amount of highly repetitive DNA in the region surrounding the megasatellite was made. Total human DNA, labeled by random prime labeling, was hybridized to *Eco*RI digested cosmid 153G8 and BAC 87B23 (figure 20). One band in 153G8, containing the T7 end, hybridized to the probe. Therefore, the region including and immediately flanking the megasatellite contains none of the highly repetitive DNA, such as *Alu* and LINES, found in high copy number in the genome. A number of bands in 87B23 hybridized to the probe. The enrichment of the BAC library for chromosome 8 sequences was carried out by hybridization of Inter-*Alu* PCR products to the total BAC library, therefore a deficit of these sequences in the region immediately surrounding the megasatellite may explain the failure to

isolate BACs containing the megasatellite. However, the region flanking the DNA contained within the cosmid may be better represented in the library, as evidenced by the large number of clones isolated when screening the Research Genetics BAC library.

Isolation of BAC 87B23 allowed placement of the type I megasatellite, based on STS content, on the YAC contig physical map. However, large clones containing the type II and type III megasatellites were not isolated. Therefore, YACs were investigated for their megasatellite content in order to place the remaining megasatellites on the 8p map.

3.7 Megasatellite Content in Chromosome 8 YACs

A minimum tiling path in YACs, across the short arm of chromosome 8, was assembled by examination of the MIT/Whitehead Institute database. FISH analysis confines the chromosome 8 megasatellites to distal 8p. Therefore, YACs spanning from the telomere region to marker D8S552 (located in 8p22) (figure 19) were examined for their megasatellite content using the characteristic restriction enzyme analysis and hybridization with the 4.7 kb megasatellite fragment (figure 21). Three YACs were found to contain the megasatellite. YAC 967c11, contains a type II megasatellite, YAC 764c7 (chimeric for chromosome 8) contains a type III megasatellite and YAC 773g4 both a type II and type III megasatellite. No YACs containing the type I megasatellite were recovered. Isolation of YACs containing the entire type III megasatellite allowed the characteristic bands to be modified to include a 4.0 kb *Eco*RI fragment. As well, *Nde*I digestion results in a single large band and *Bss*HII digestion results in two large bands (not resolved in figure 21) and no 4.7 kb band (table 6). Four other YACs (737e5, 809h8, 871f3, 920d12), known to contain the markers D8S1619, D8S1993 and D8S1640 found in BAC 87B23 (which contains a type I megasatellite) were analyzed for megasatellite sequences (data

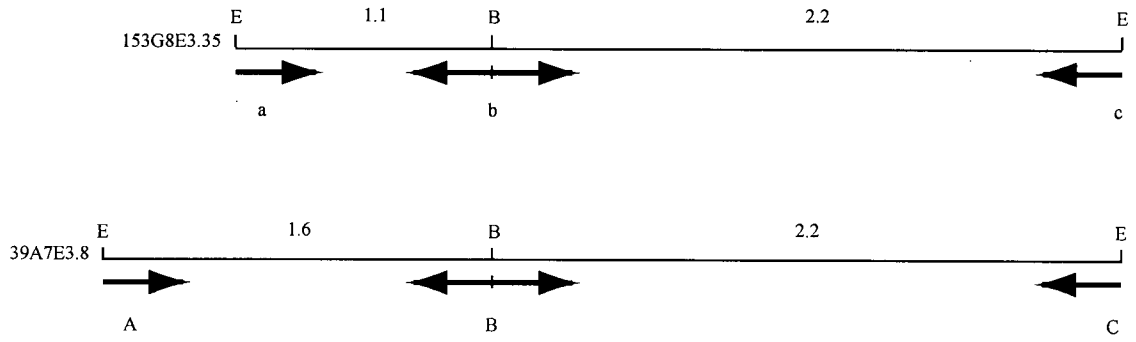


Figure 18: Schematic representation of sequenced fragments. 153G8E3.35 and 39A7E3.8 are the 5' flanking fragments of the type I and type II megasatellites respectively. Sequences are represented by arrowheads, pointing the direction of sequencing. Sequences are: a. 153G8BE1.1-M13R, b. 153G8BE1.1-M13F+153G8BE2.2-M13F, c. 153G8BE2.2-M13R, A. 39A7BE1.6-M13R, B. 39A7BE1.6-M13F+39A7BE2.2-M13F, C. 39A7BE2.2-M13R. All sequences share some homology with the 4p megasatellite sequence. Sequence a is highly homologous to A, sequence b is highly homologous to B, sequence c is highly homologous to C (see Appendix 1).

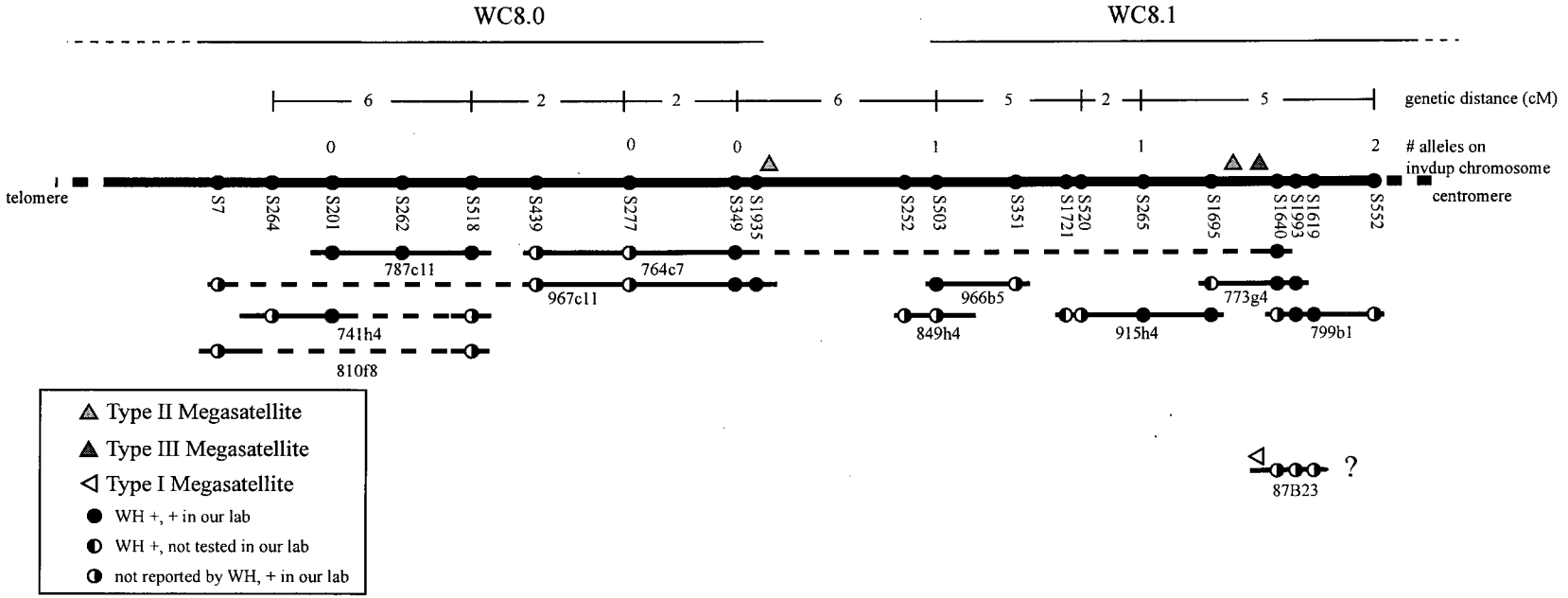


Figure 19: Localization of megasatellites by YAC/BAC content analysis. Data collected by the MIT/Whitehead Institute (WH) and in our laboratory. WC8.0 and WC8.1 are YAC contigs assembled at the MIT/Whitehead Institute, modifications to STS marker order have been made, additional STS markers have been tested. Dotted lines represent gaps in YACs. All clones, with the exception of BAC 87B23, are YACs. Orientation of megasatellites on chromosome 8 is not known. Location of BAC 87B23 is based on STS content of markers placed on the MIT/Whitehead YAC map. These markers (D8S1640, D8S1993, D8S1619) may be duplicated on chromosome 8, with the duplicated copies located within the gap between the YAC contigs.

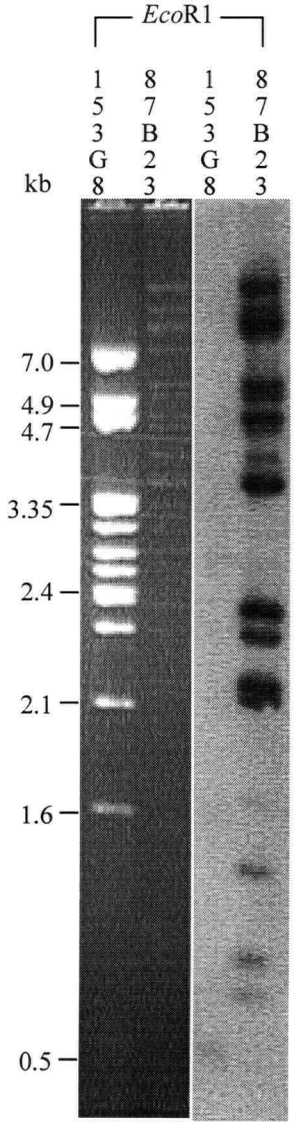


Figure 20: Hybridization of total human DNA to clones containing the megasatellite. Cosmid 153G8 and BAC 87B23 were digested with *Eco*RI. The *Eco*RI restriction fragment pattern is shown on the left, hybridization of total human DNA is shown on the right.

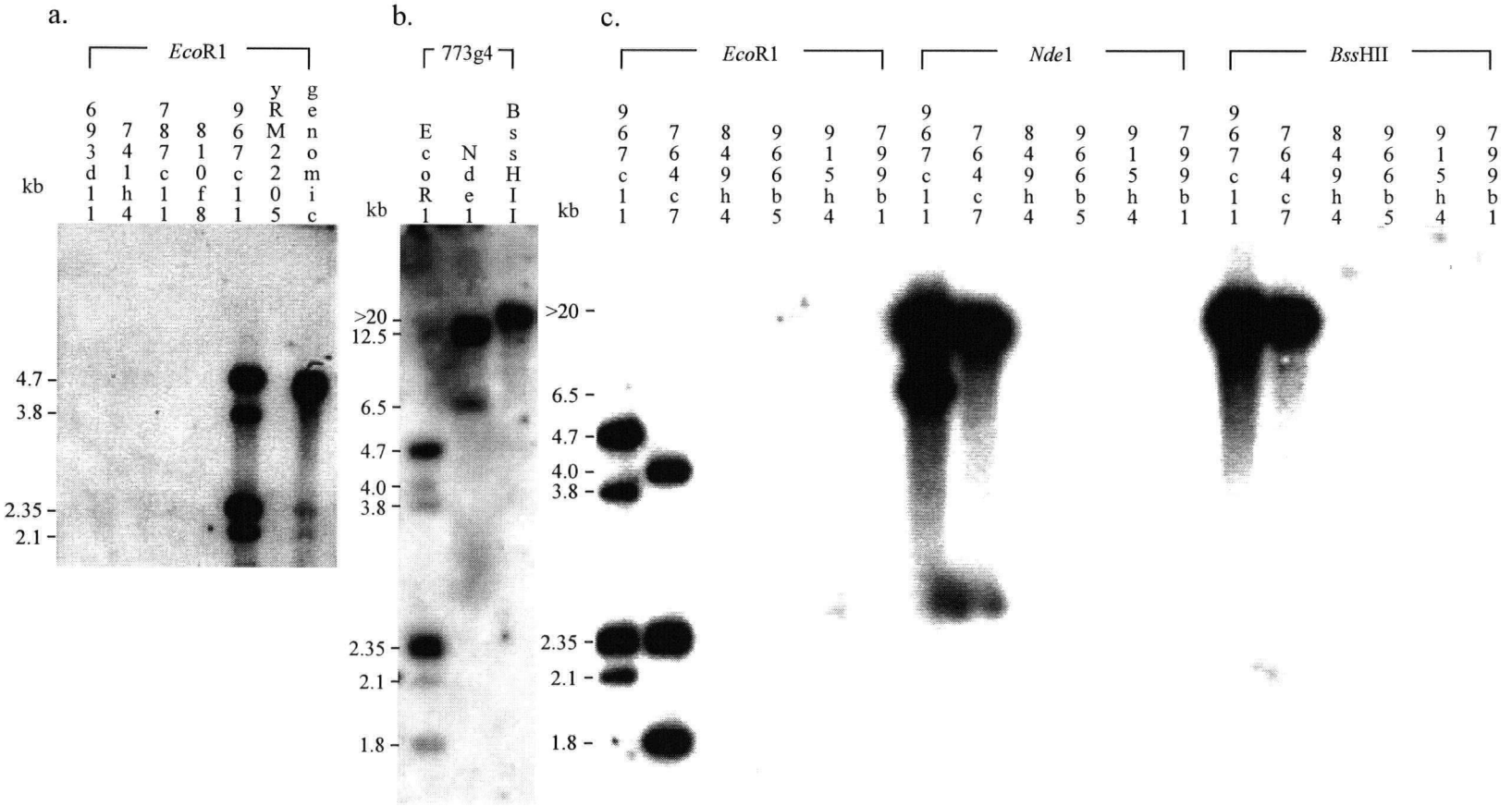


Figure 21: Analysis of YAC clones by hybridization of megasatellite sequence. YACs were digested with the enzymes that characterize the megasatellite types. a. hybridization of megasatellite to YACs from distal 8p, digested with *Eco*RI, b. hybridization of megasatellite to YAC 773g4 digested with the characterizing enzymes, c. hybridization of megasatellite to YACs digested with the characterizing enzymes.

not shown), no type I megasatellite was recovered in these YACs. The MIT/Whitehead database maps YAC 967c11 and YAC 773g4 to two different locations on chromosome 8, based on STS content (figure 19), lending further evidence for the existence of more than one location of the megasatellite on 8p. The MIT/Whitehead Institute database reports a 6% false positive rate in their data (Hudson *et al.*, 1995). Therefore, the STS content of these YACs must be confirmed.

3.8 STS Content Analysis of Megasatellite Containing Clones

STS content analysis of clones containing megasatellite sequences is important for accurately placing the megasatellites on chromosome 8p. PCR analysis has been carried out in our laboratory on the YAC clones for the STS markers D8S1640, D8S1993, D8S1619, D8S1819, and D8S1935 (with the exception of D8S1819 and D8S1935 data presented in section 3.2, data not shown). However, analysis of these clones, and the BAC 87B23, by hybridization of the PCR products of these markers more accurately confirms STS content than does PCR.

A partial PCR product of D8S1819 (same $(CA)_n$ as D8S349), constructed by digestion of the PCR product with *Hae*III to remove the $(CA)_n$, was used as a hybridization probe to clones containing the megasatellite; BAC 87B23 (type I), and the YACs 967c11 (type II), 764c7 (type III), and 773g4 (types II and III). Its presence in YACs 967c11 and 764c7, and its absence in BAC 87B23 and YAC 773g4, was confirmed (figure 22a). The PCR product of D8S1935 was used as a hybridization probe to the same clones. Its presence in YAC 967c11, and absence in BAC 87B23 and YACs 764c7 and 773g4, was confirmed (figure 22b).

BAC 87B23 contains the markers D8S1640, D8S1993, D8S1619, and a type I megasatellite. The content of these STS markers in YACs containing megasatellite sequences was investigated. The PCR product of each of the three markers (primer sequences and annealing temperatures

listed in table 2) was hybridized to *Eco*RI and *Hind*III restriction enzyme digests of clones containing the megasatellite; YACs 764c7 (type III), 967c11 (type II), and 773g4 (type II and type III), and BAC 87B23 (type I) (figure 23). YACs 764c7 and 773g4 contain D8S1640, and 773g4 contains D8S1993. YAC 967c11 does not contain the three markers. The fragment sizes seen in the YACs were consistent with those seen in 87B23.

Unexpectedly, D8S1640 hybridizes to two *Hind*III bands in YAC 764c7. As well, D8S1993 hybridizes to two *Eco*RI bands in YAC 773g4. However, the sequences reported for the PCR products D8S1640 and D8S1993 do not contain either an *Eco*RI or a *Hind*III restriction enzyme cleavage site. These data suggest that either these sequences are present more than once in the clones or, a restriction enzyme recognition site has been introduced into the sequence. Regardless of the correct interpretation, these data suggest that large reiterated sequence may extend to include the markers D8S1640, and D8S1993. Investigations were undertaken to determine the size of the LRS in which each class of megasatellite is embedded.

3.9 Analysis of the LRS on Chromosome 8p

Additional experiments were carried out on the BAC 87B23 and the YACs 967c11, 764c7, 773g4, and 799b1 in order to determine the size of the LRS. YAC 799b1, which does not contain a megasatellite, was included in the experiments because it contains D8S1640, D8S1993, D8S1619, and D8S552, placing it proximal to the megasatellite on the YAC contig map (figure 19).

BAC 87B23 contains the markers D8S1640 and D8S1993, which may be included in the LRS. In order to compare homology between YACs, hybridization probes were prepared from the insert of BAC 87B23, isolated from a pulsed field gel, and the SP6 and T7 ends of 87B23,

isolated by bubble PCR. BAC 87B23 and YACs 967c11, 764c7, 773g4, and 799b1 were digested with *Sst*I, separated on an agarose gel, transferred to nylon membrane and hybridized with the insert of BAC 87B23, the T7 end of 87B23, and the SP6 end of 87B23, respectively. The results of these experiments are presented in figures 24 and 25, and summarized here. YAC 799b1 does not hybridize with either end of 87B23 (data not shown) but cross hybridizes with the insert of 87B23. However, the majority of these bands are not the same size as those present in the other YACs (figure 24, bands are faint, all ~6 kb and larger). YAC 764c7, a known chimera, contains the SP6 end of 87B23 in a 9.7 kb *Sst*I fragment, and contains a number of *Sst*I fragments that cross hybridize with the insert of 87B23 with similar in size to those in YACs 967c11 and 773g4. Most surprisingly, YACs 773g4 and 967c11 have almost identical hybridization patterns when hybridized with the insert of 87B23 (figure 24), and this pattern differs from the *Sst*I pattern of BAC 87B23. These YACs map to different regions of 8p using STS content (figure 19), therefore, the similarity in hybridization pattern must represent homology of these regions and not true overlap. The T7 end of 87B23 hybridizes to two *Sst*I fragments in 773g4, one 20 kb, the other just slightly larger than 20 kb (figure 25). YAC 967c11 does not hybridize with either the T7 or the SP6 end of 87B23. The similarity in the hybridization pattern of YACs 967c11 and 773g4 using BAC 87B23 for hybridization could mean that the LRS is not entirely contained within BAC 87B23. Consequently, the ends of BAC 87B23 were used to isolate walk clones.

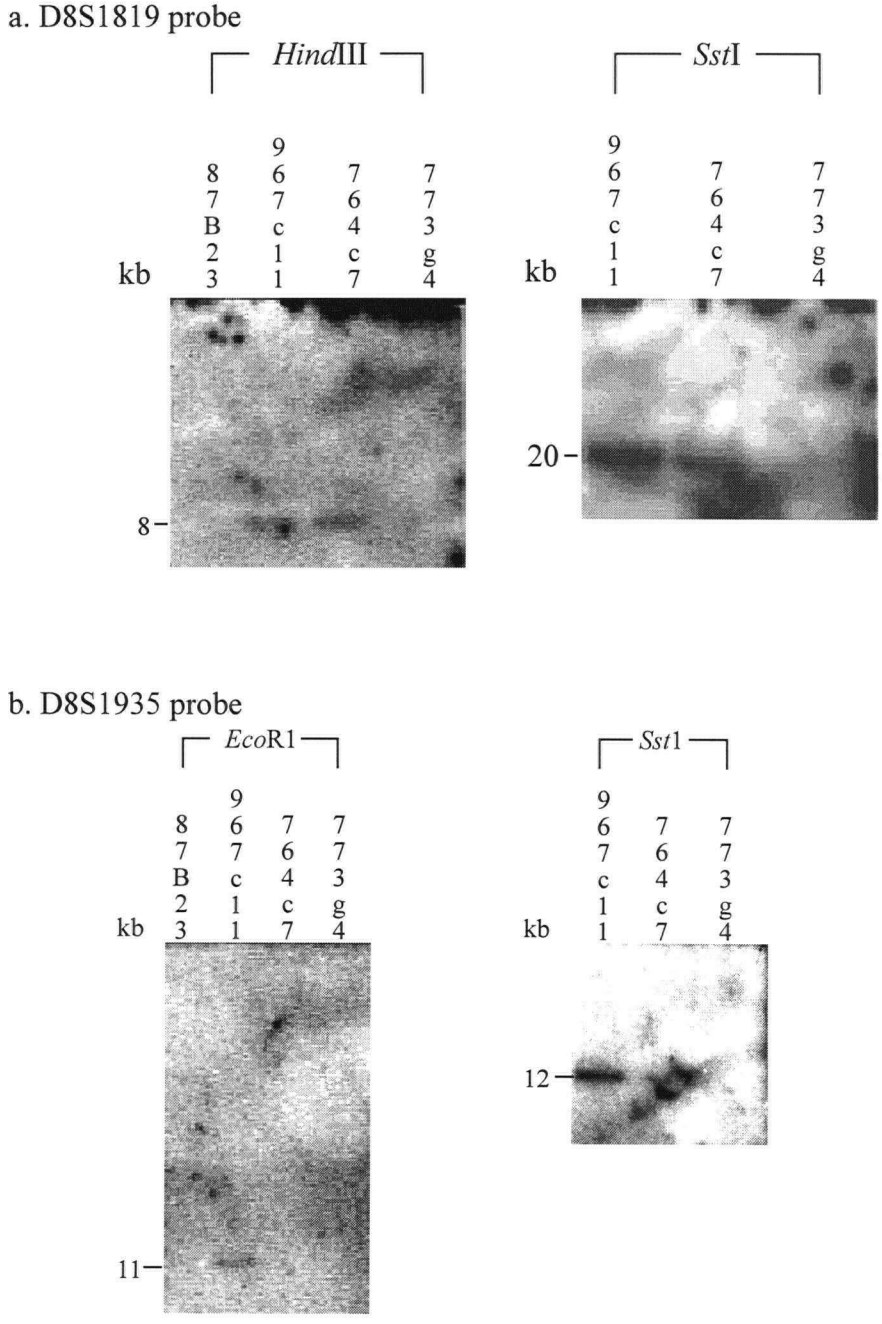


Figure 22: D8S1819 and D8S1935 content analysis by hybridization. With the exception of BAC 87B23, all clones are YACs. D8S1819 is a polymorphic STS marker containing a (CA)n. The (CA)n has been removed by digestion and recovery of flanking sequence. D8S1935 is a non-polymorphic STS marker. a. Clones digested with *Hind*III, or *Sst*I, and hybridized with the D8S1819 probe. b. Clones digested with *Eco*RI, or *Sst*I, and hybridized with the D8S1935 probe.

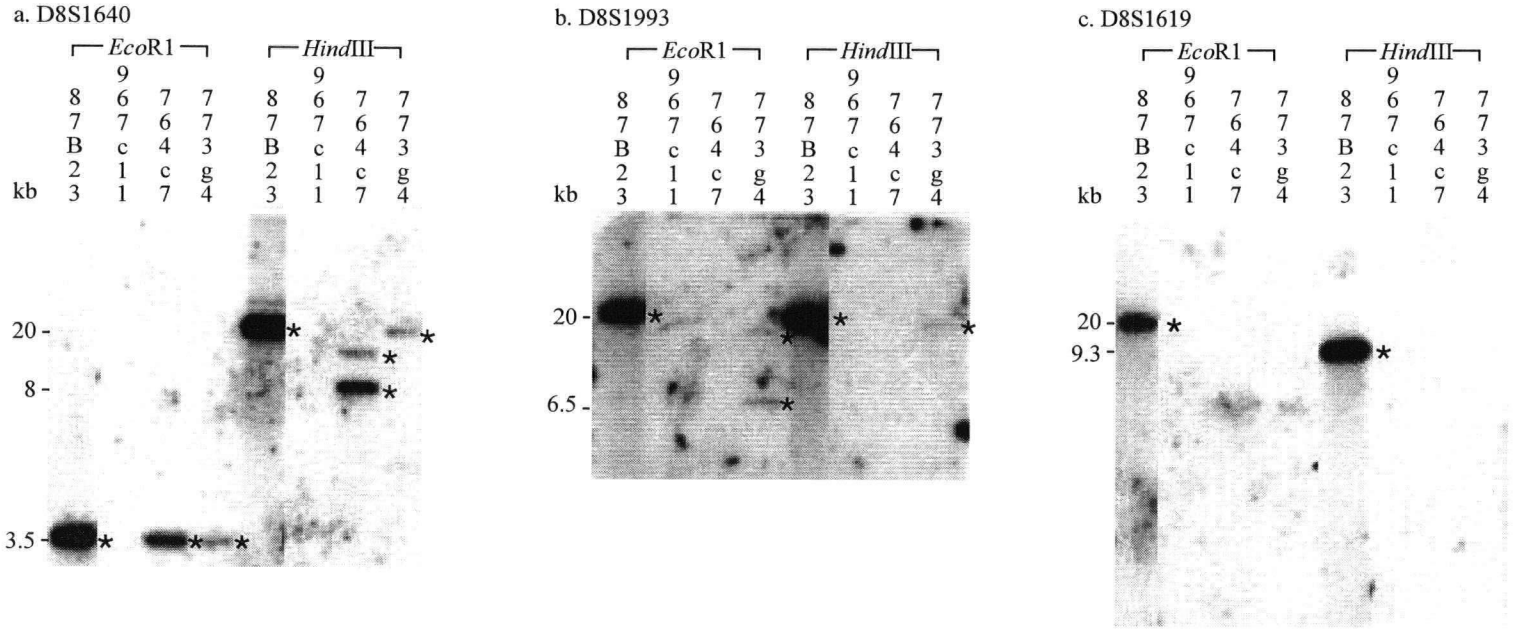


Figure 23: Analysis of STS content of YACs and BAC by hybridization. With the exception of BAC 87B23, all clones are YACs. An asterick is located to the left of each hybridized DNA fragment. Clones were digested with *Eco*RI or *Hind*III and hybridized with: a. the PCR product of D8S1640, b. the PCR product of D8S1993, c. the PCR product of D8S1619. Extraneous lanes have been deleted from the autoradiographs.

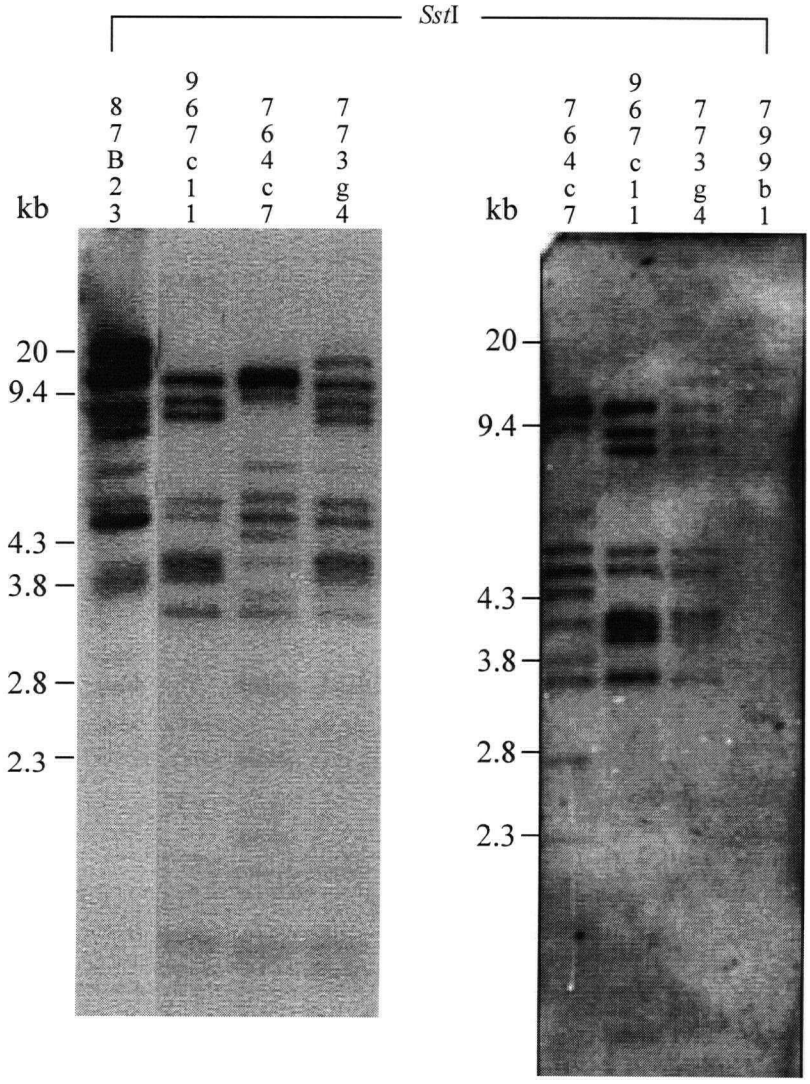


Figure 24: Analysis of homology between YACs by hybridization of BAC 87B23. Clones are digested with *Sst*I. With the exception of 87B23, all clones are YACs. The autoradiograph on the left is a composite of two different exposure times.

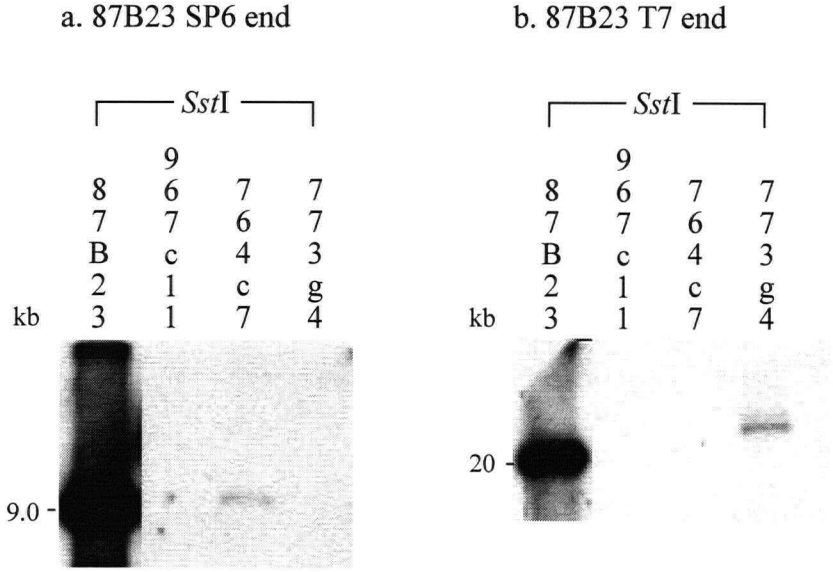


Figure 25: Analysis of YACs by hybridization of BAC 87B23 T7 and SP6 ends.
Clones were digested with *Sst*I and hybridized with a. 87B23 SP6 end, b. 87B23 T7 end.
Extraneous lanes have been deleted from the autoradiographs.

3.9.2 Chromosome Walking from BAC 87B23

The cosmid and enriched chromosome 8 BAC library were screened with the T7 and SP6 ends of 87B23 to isolate clones for hybridization against the YACs.

The T7 end of 87B23 hybridized to cosmids 113F7, 172D7 and 128H8 (figure 26c) that were previously isolated (figure 15, 16). Cosmids 172D7 and 128H8 contain a type III megasatellite, cosmid 113F7 contains either a type I or a type II megasatellite. All three cosmids contain 3′ sequences flanking the megasatellite, but no 5′ flanking sequences. As well, only ~9 kb of DNA is common to these 3 cosmids, but not common to the other cosmids containing megasatellite sequences (figure 15). Therefore, the T7 end of 87B23 is located within the LRS 3′ region flanking the megasatellite, and does not represent the true end point of the LRS. The distance from the T7 end of 87B23 to the megasatellite is estimated from figure 15 to be approximately 20 kb. The SP6 end of 87B23 hybridized to cosmid 128H10. This cosmid was not investigated further because the T7 end was anchored within 20 kb of the megasatellite.

The T7 end hybridized to two BACs, while the SP6 end hybridized to four BACs, isolated from the 8p enriched BAC library. Three of these BACs, 223B23, 67B6 and 286D2, were chosen for further analysis (figure 27). All three have *Sst*I patterns that differ from the *Sst*I pattern of 87B23 (figure 27a). BACs 67B6 and 223B23 hybridized to the T7 end of 87B23, BAC 286D2 to the SP6 end of 87B2. BAC 223B23, almost completely cross hybridizes with BAC 87B23. BAC 67B6 cross hybridizes to BAC 87B23, however, by summing the sizes of the restrictions fragments that do not cross hybridize, BAC 67B6 is estimated to extend 20 kb past the T7 end of BAC 87B23 (figure 27a, b). BAC 286D2 cross hybridizes to BAC 87B23,

however, by summing the sizes of the restriction fragments that do not cross hybridize, BAC 286D2 is estimated to extend 80 kb past the SP6 end of 87B23 (figure 27 a, b).

The human DNA inserts of 223B23, 67B6, and a partial insert of 286D2 including the SP6 end of 87B23 and the ~80 kb region extends past BAC 87B23, were isolated by PFGE. Their respective sizes are 130 kb, 97 kb, and 80 kb (data not shown). Each insert was used as a hybridization probe to *Sst*I digests of BAC 87B23, and YACs 764c7, 967c11, and 773g4 (figure 28) and compared to the hybridization patterns of BAC 87B23 (figure 24). BAC 223B23 hybridizes to all but the 4.3 kb fragment hybridized by 87B23 and does not hybridize to any new fragments. BAC 67B6 does not hybridize to novel bands in the YACs. The partial insert of BAC 286D2 hybridizes to the fragment of 87B23 containing the SP6 end but does not hybridize to the YACs. This suggests that the LRS ends near the SP6 end of BAC 87B23 and may end just past the T7 end of 87B23.

The insert of BAC 223B23 is almost completely homologous with BAC 87B23, including the T7 end of 87B23 which is estimated to be within ~20 kb of the megasatellite, but the *Sst*I restriction fragment patterns differ (figure 27a). These data suggest that BAC 223B23 may contain a megasatellite differing from the type I megasatellite in BAC 87B23.

3.10 Lack of Megasatellite in BACs 223B23, 67B6, and 286D2

BAC 223B23 was analyzed for its megasatellite content by digesting with restriction enzymes *Eco*RI, *Nde*I, and *Bss*HII hybridizing with the 4.7 kb megasatellite fragment (data not shown). BAC 223B23 does not contain a megasatellite. As well, hybridization of the 4.7 kb megasatellite fragment to *Sst*I digests of BACs, including 223B23, 67B6 and 286D2, showed that none of the BACs homologous with 87B23 contain megasatellite sequences (figure 29). However, 223B23

has almost complete homology to 87B23 (only the 4.3 kb fragment, containing megasatellite sequences, does not hybridize with 223B23), is 130 kb in size, and overlaps with the T7 end of 87B23, which, based on cosmid data, is approximately 20 kb away from the megasatellite. The lack of a megasatellite in BAC223B23, and the isolation of more than 3,000 BAC clones in the library screen of the Research Genetics, Inc. BAC library with the T7 and SP6 ends of 87B23, raises the question of whether these sequences originate from 8p. This was addressed by FISH.

3.11 Mapping the LRS by FISH using BAC 223B23

BAC 223B23 is estimated to contain approximately 110 kb of the 5′ flanking sequence and approximately 20 kb of the 3′ flanking sequence of the megasatellite (figure 27), and is therefore a useful FISH probe for the LRS. FISH of the biotin-labeled insert of 223B23 to normal control chromosomes resulted in a strong hybridization signal at 8p23, analogous to the location of the megasatellite, and weaker signals on many other chromosomes (figure 30).

3.12 Characterization of Five Additional Patients

A patient with what appeared to be a standard inv dup(8p) chromosome, three patients with unusual inv dup(8p) chromosomes, and the patient of Dhooge *et al.* (1994) were analyzed by STS genotyping and FISH analysis.

3.12.1 Polymorphic STS Genotyping in Five Patients

STS genotyping at D8S201, D8S349, D8S503, D8S265, D8S552, and D8S135 (figure 3, table 2), was carried out on these five patients (patients 10-14) and their parents (or sibling, in the case of patient 14) to allow comparison between these patients, and previously reported inv dup(8p) patients. Markers D8S201 and D8S349 are located to the region of 8p deleted from the

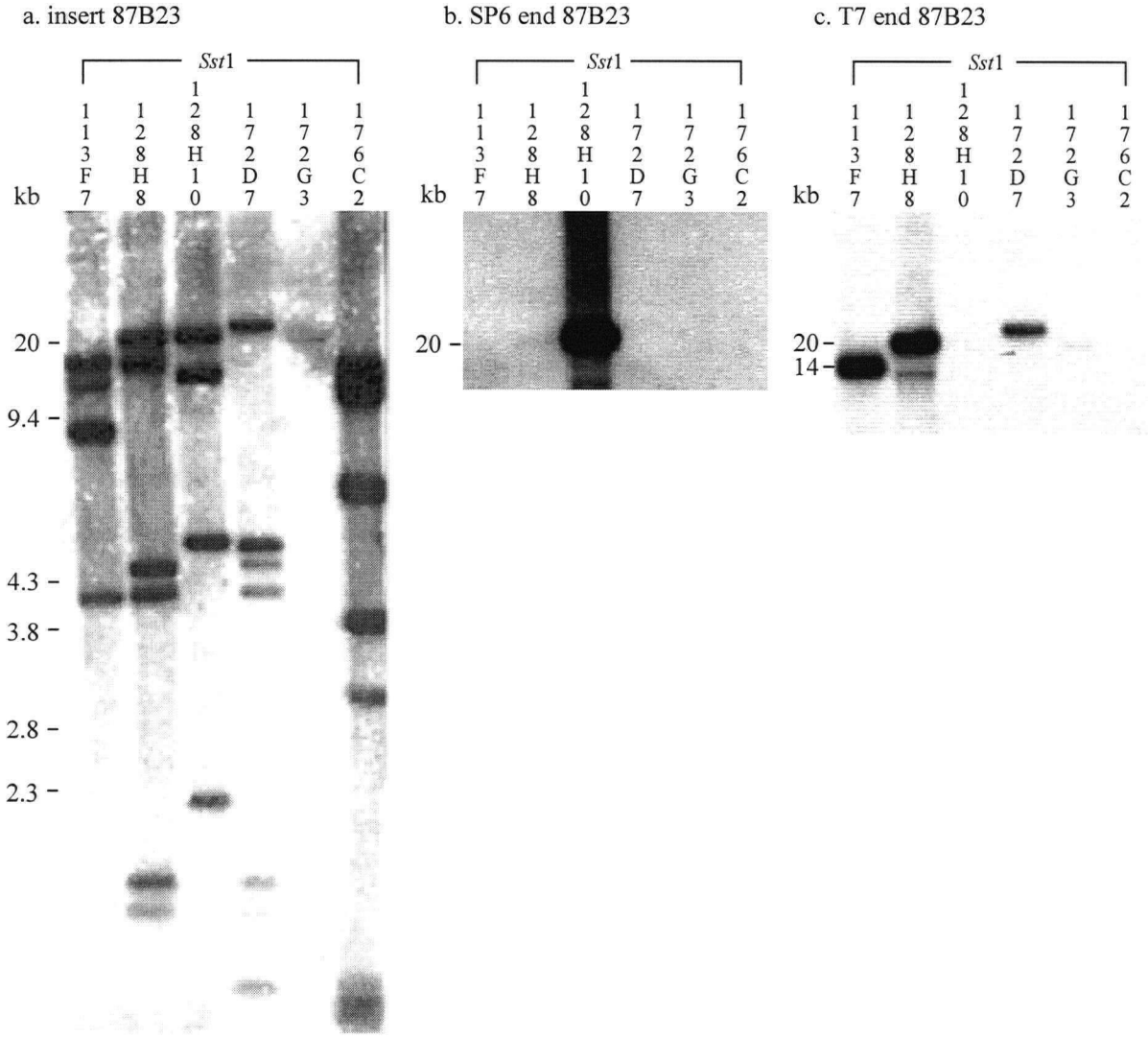


Figure 26: Analysis of cosmid clones isolated by chromosome walking from BAC 87B23. The T7 and SP6 ends of BAC 87B23 were used to screen the chromosome 8 specific cosmid library. Cosmid clones were digested with *Sst*I and hybridized with: a. the insert of 87B23, b. the SP6 end of 87B23, c. the T7 end of 87B23.

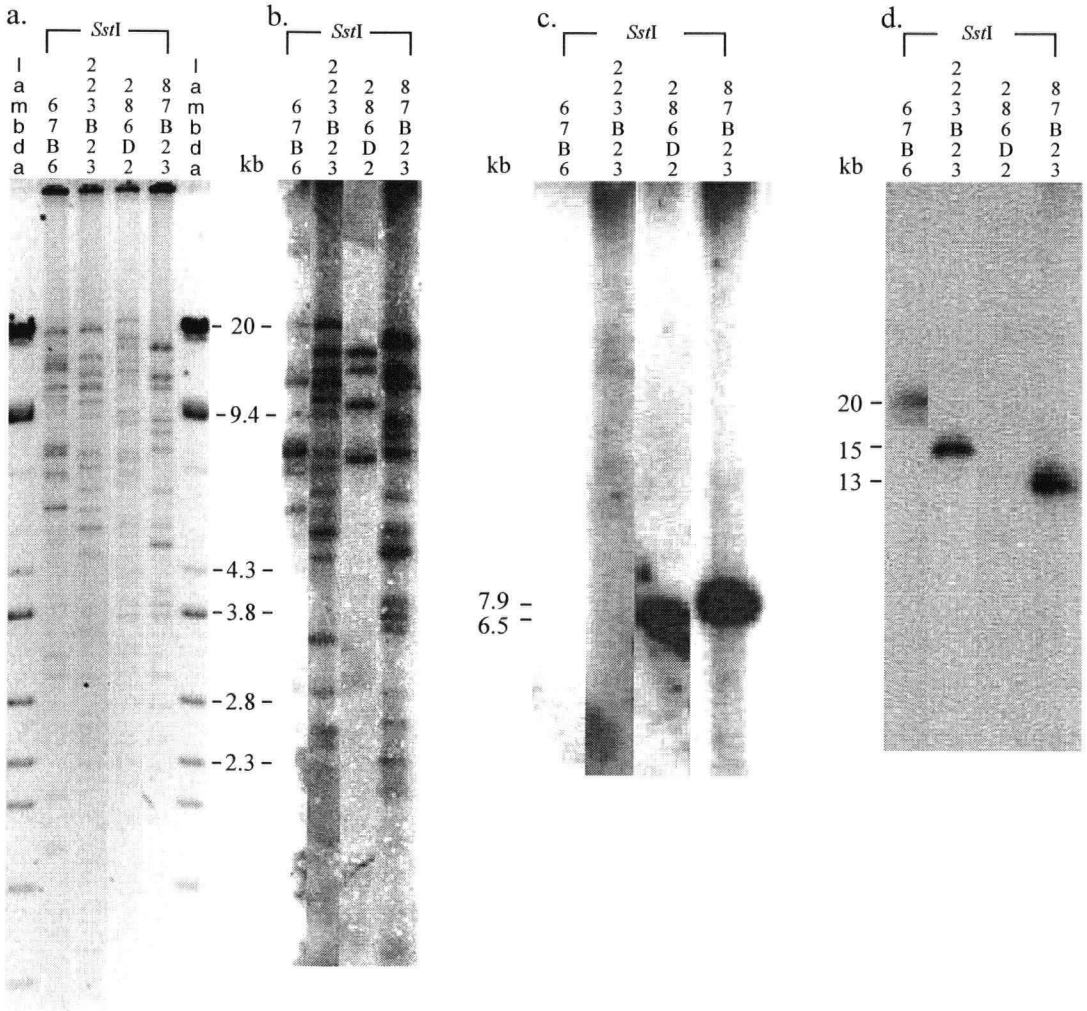


Figure 27: Analysis of BAC clones isolated by chromosome walking from BAC 87B23. The T7 and SP6 ends of 87B23 were used to screen the chromosome 8 enriched BAC library. a. Negative image of BAC *Sst*I digestion, lambda DNA is digested with *Hind*III/*Sst*II, b. Hybridization of the insert of BAC 87B23 to "a", c. Hybridization of the SP6 end of 87B23 to "a", d. Hybridization of the T7 end of 87B23 to "a". Extraneous lanes have been deleted from the autoradiographs.

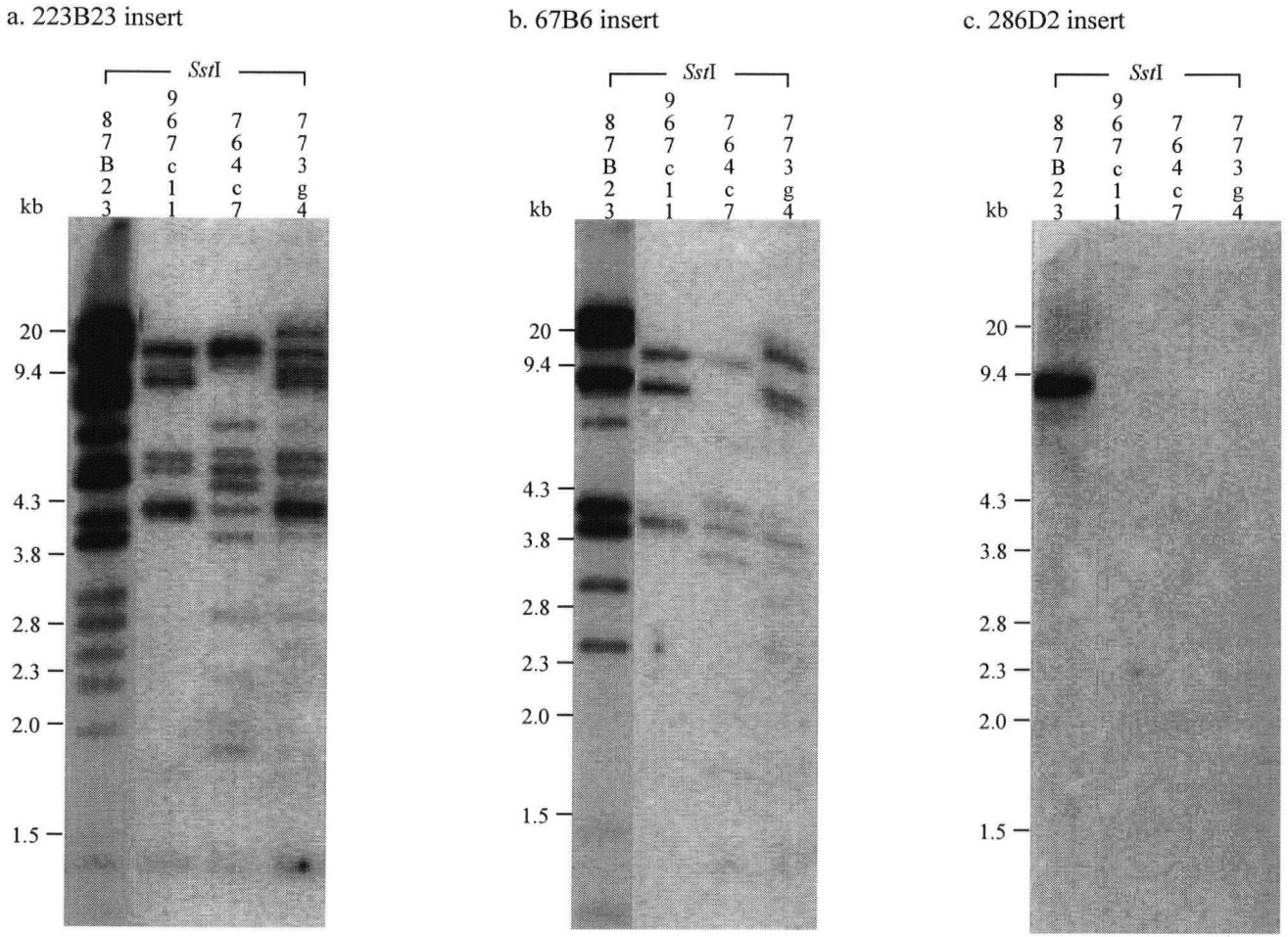


Figure 28: Analysis of YACs and BAC by hybridization of BAC clones isolated by chromosome walking.
With the exception of BAC 87B23, all digested clones are YACs. Clones were digested with *Sst*I and hybridized with:
a. BAC 223B23, b. BAC 67B6, c. BAC 286D2. Autoradiographs are composites of two different exposure times.

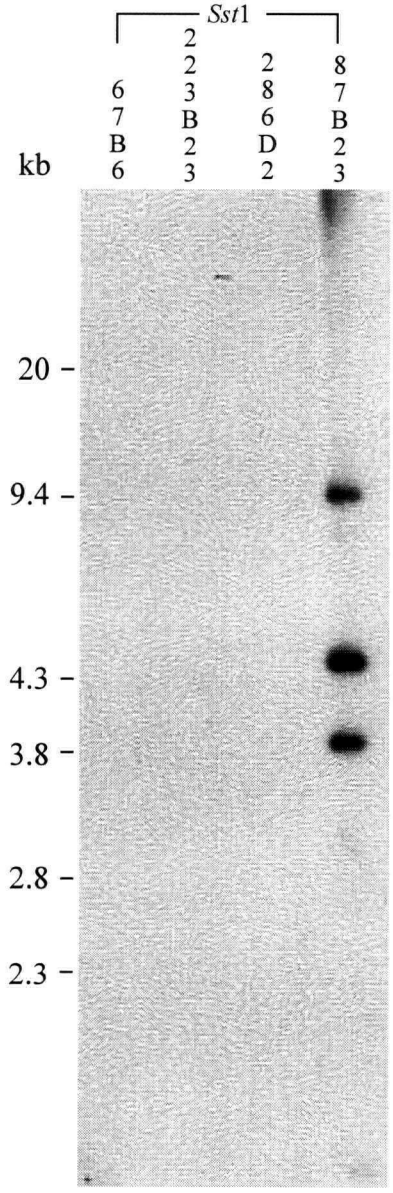


Figure 29: Analysis of megasatellite content of BACs isolated by chromosome walking. Clones were digested with I and hybridized with 153G8E4.7 and 39A7E4.7 which contain the complete megasatellite sequence.

Figure 30: FISH analysis with BAC 223B23. The insert of BAC 223B23 and a chromosome 8 specific centromere probe were hybridized to normal metaphase chromosomes. Hybridization occurred to 8p23 (large arrowheads), at the location analogous to the hybridization location of the megasatellite, as well as to number of other locations throughout the genome.

inv dup(8p) chromosome. D8S349 is deleted in 11/11 informative patients (Floridia et al., 1996) and represents the most proximal marker in the region deleted from the inv dup(8p) chromosome. Markers D8S503 and D8S265 are localized to the single copy region at the center of symmetry of inv dup(8p) chromosomes (Floridia *et al.*, 1996, and table 3). D8S552 is duplicated on the inv dup(8p) chromosome in 9/9 informative patients and delimits the distal extent of duplication in those patients who have a single copy region (Floridia *et al.*, 1996). Marker D8S135 is located in band 8p11.22 and may be present in one or two copies on the inv dup(8p) chromosome, depending on the extent of chromosomal DNA included in the duplication. Examples of raw genotyping data are presented in figure 31. Genotyping results are summarized in table 8. Patient information, not reported elsewhere, and a discussion of genotyping results, is presented below.

Interpretations of genotyping results were made in the following way: In all cases, at a given locus, a single allele was considered to be present on the normal chromosome of the patient. Then, based on how informative the genotyping was, the number of alleles on the aberrant chromosome could be interpreted as: zero (deleted), one (single copy), two (duplicated), zero/one, one/two or uninformative. Inheritance of a single allele at a given locus was considered evidence that the marker was deleted from the aberrant chromosome and that the chromosome was derived from the parent whose allele was not inherited. The inheritance of three alleles at a given locus was considered evidence that the marker was duplicated on the aberrant chromosome and that the chromosome was derived from the parent from whom two alleles had been inherited. The inheritance of two alleles from one parent was considered evidence that the aberrant chromosome was derived from an interchromosomal event.

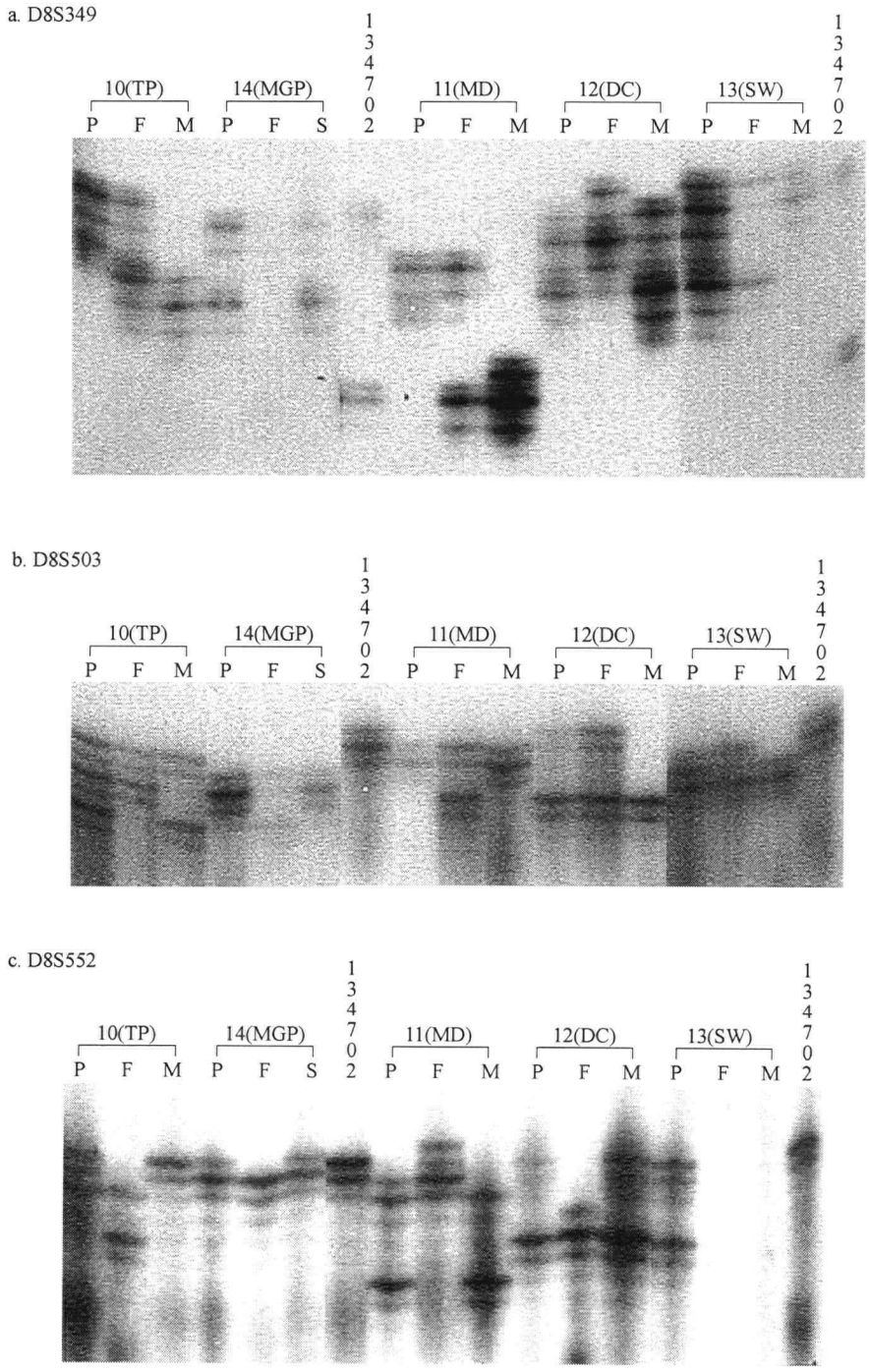


Figure 31: Examples of genotyping of patients 10-14. a. STS polymorphic marker D8S349, b. STS polymorphic marker D8S503, c. STS polymorphic marker D8S552. Patient numbers and initials are shown above each triad. P=patient, F=father, M=mother, S=sister, 134702 is a *CEPH* family member included as a positive control. Extraneous lanes have been deleted from the autoradiographs.

Table 8: Summary of polymorphic STS genotyping data for patients 10-14.

Markers are listed in order from telomere to centromere on the normal chromosome. Within each triad, allele sizes are reported as A-D, A being smallest, D being largest. Genotypes are presented in the order patient, father, mother or, for patient 14, patient, father, sister. The dashed line indicates failure of amplification of alleles.

	10 TP			14 MGP			11 MD			12 DC			13 SW		
D8S201	B	B	AC	AC	AB	AC	C	BC	A	AD	AB	CD	BC	AC	B
D8S349	C	BC	AB	BC	AC	BC	C	AC	AB	AB	BD	AC	ABC	AC	BC
D8S503	ABC	BC	AC	BC	AC	BC	BC	AC	BC	AB	AB	A	ABC	AC	AB
D8S265	BCD	AD	BC	AB	AB	AB	BC	AC	B	AB	B	AB	B	B	AB
D8S552	BC	AB	C	AB	A	AB	ABC	CD	AB	AC	AB	AC	AB	---	---
D8S135	A	AB	AB	A	A	A	A	A	A	B	AB	BC	AB	AC	BC

3.12.1 Patient 10 (TP)

Patient 10 (TP) was born to karyotypically normal parents. As a child she presented with developmental delay, abnormal facies, and behavioural problems. At 29, she is severely mentally retarded, has a large mouth, large skull and upturned nose, and has a vocabulary of 10 words. Karyotype analysis revealed an inverted duplication of 8p: 46, XX, inv dup(8)(p12→p23.1) de novo. Genotyping at distal markers D8S201 and D8S349 is consistent with lack of maternal alleles on the inv dup(8p) chromosome (table 8). Genotyping at markers D8S503 and D8S265 is consistent with inheritance of two different maternal alleles at these loci (table 8). Genotyping at D8S552 is consistent with inheritance of at least one maternal allele but is uninformative for copy number (table 8). It likely D8S552 is duplicated since markers distal to this locus are duplicated and the karyotype is consistent with duplication of the region to which D8S552 has been localized. A single maternal allele is inherited at D8S135 indicating that the duplication does not extend to include this locus. The absence of maternal alleles at D8S201 and D8S349, and the inheritance of two different maternal alleles at D8S503 and D8S265, is consistent with the inversion duplication being derived from a maternal inter-chromosomal event. With this interpretation, genotyping is consistent with paternal alleles being inherited in one copy at all informative markers.

3.12.2 Patient 11 (MD)

Patient 11 (MD) presented with developmental delay, hypotonia, and an unusual face at approximately 1 year. She is 3 years old, at the time of this report. Cytogenetic analysis suggests that MD carries a classic inv dup(8p) chromosome with a center of symmetry at 8p23.1,

but the with the aberrant 8p telomere capped by satellites. Parental chromosomes are normal. A number of experiments were carried out in a diagnostic laboratory: FISH using a whole chromosome 8 paint showed that the entire aberrant chromosome, except the satellites, originated from chromosome 8; the distal p arm of the aberrant chromosome stains with silver staining specific to the nucleolar organizer region (NOR); FISH using the acrocentric α-satellite centromere probes (13, 14,15, 21, 22 (ONCOR)) showed no evidence of these sequences near the telomere of the aberrant chromosome. Genotyping of this patient at the above mentioned STS markers was consistent with a classic inv dup(8p) patient (table 8). Lack of inheritance of maternal alleles is seen at D8S349 and D8S201. At least one maternal allele is inherited at D8S503, and, at D8S265. Two different maternal alleles have been inherited at D8S552. D8S135 is uninformative. These data are consistent with maternal origin as a result of an interchromosomal event. The origin of the satellites is unknown but they are presumably derived from maternal chromosomes. With the interpretation that the aberrant chromosome is of maternal origin, a single paternal allele is inherited at all informative markers.

3.12.3 Patient 12 (DC)

Patient 12 (DC), who is 6 years of age at the time of report, was investigated due to hypotonia as an infant. She has the mental capacity of a 3-4 year old, speaks 50 words, can walk, and is still in diapers. She has a large skull and ears, and congenital heart anomaly. Her karyotype is inv dup(8)(p12→p23.?2). She has a large G band located at the center of symmetry. Parental chromosomes were normal. FISH analysis carried out in a diagnostic laboratory with a whole chromosome 8 paint indicates that the duplicated material is from chromosome 8. Genotyping at D8S201 and D8S349 is informative showing the inheritance of a single maternal

allele and a single paternal allele (table 8). These STS markers are normally deleted on the inv dup(8p) chromosome. However, the presence of these markers on the aberrant chromosome is consistent with the cytogenetic evidence for the presence of band p23.2. Markers D8S503, D8S265 and D8S552 are informative for the inheritance of at least one maternal and one paternal allele, but not for copy number. D8S135 is uninformative for copy number. From the genotyping results alone, the presence of a duplication could not be detected and the origin of the aberrant chromosome could not be determined. Whether this is a result of uninformative markers, or, whether the aberrant chromosome arose as a result of an intrachromosomal event is unknown. DC carries an inversion duplication of a different type to those previously reported.

3.12.4 Patient 13 (SW)

Patient 13 (SW) was ascertained with amenorrhea at 17 years of age. She has had some difficulties in school and is therefore thought to be mildly mentally retarded. She has a large skull but no striking dysmorphism. Analysis revealed a karyotype of inv dup(8)(p21.3→p23.?2). There is a G band at the center of symmetry of the aberrant chromosome. However, this band is not as large as that seen in patient DC. Parental karyotypes were normal. Genotyping is consistent with the inheritance of at least one maternal allele, and a single paternal allele, at D8S201 (table 8). Three alleles are inherited at D8S349 and at D8S503. Whether the additional alleles were inherited from the mother or the father could not be determined. Parental alleles at D8S552 failed to amplify. However, the patient is heterozygous. D8S135 is informative for the inheritance a single maternal and a single paternal allele. The origin of the aberrant chromosome could not be determined.

3.12.5 Patient 14 (MGP)

Patient 14 (MGP) was previously reported by Dhooge *et al.* (1994). The aberrant chromosome was inherited from his mother, who carries the same chromosome. FISH analysis is consistent with the extra material being of chromosome 8 origin, however, orientation and whether it originated from bands p22→p23.1 or from p21.3→p22, is unknown (Dhooge *et al.*, 1994). Maternal DNA was unavailable for genotyping, however, DNA from his sister who carries the same aberrant chromosome was genotyped. Genotyping of STS markers D8S201, D8S349, D8S503, D8S265, D8S552, and D8S135 was consistent with the inheritance of at least a single maternal allele and a single paternal allele (table 8). A third allele was not present at any of the loci.

3.13 Mapping the Megasatellite to Patient Chromosomes Using FISH

The hypothesis for the formation of inv dup(8p) chromosomes predicts that the location of repetitive elements will not change when compared to the normal chromosome. A cosmid, 153G8, containing the megasatellite sequences (figure 13) was chosen for FISH analysis of patient metaphase chromosomes.

Cosmid 153G8 was labeled with biotin by nick translation and used as a hybridization probe, along with a chromosome 8 specific centromere probe (Oncor), to patient metaphase chromosomes (FISH analysis kindly provided by S. Jurenka). Due to the lack of highly repetitive DNA in the cosmid (see section 3.6), preannealing was not required. Patients GS, a 'classic' inversion duplication (8p) patient (Dill *et al.*, 1987), TP, an inversion duplication (8p) patient lacking the single copy region, and MGP, either an inversion or direct duplication patient (Dhooge *et al.*, 1994), provided target chromosomes. As can be seen in figure 32 the

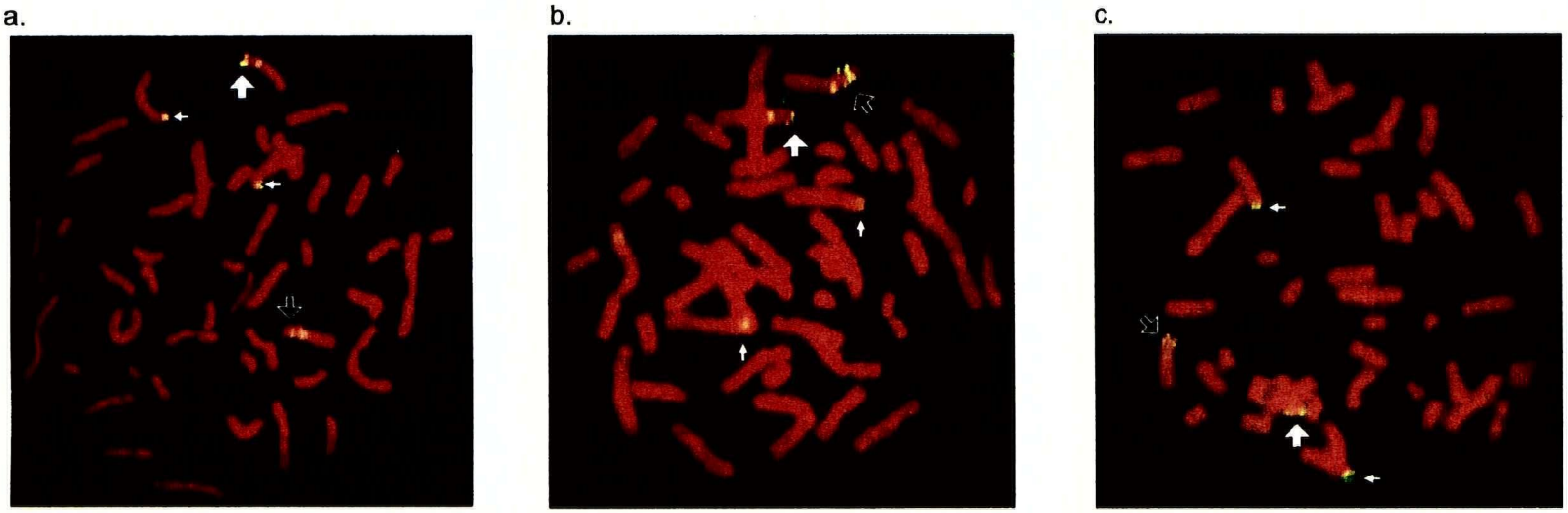


Figure 32: Hybridization of megasatellite and chromosome 8 specific centromere probe to patient metaphase chromosomes. In all cases, in addition to the chromosome 8 centromere probe, a single signal is seen on the chromosomes 4 (small arrowheads) and on the normal chromosome 8 (large solid arrowhead). a. patient GS, a single signal is seen on the aberrant chromosome (open arrowhead), b. patient TP, a diffuse signal is seen on the aberrant chromosome (open arrowhead), c. patient MGP, the centromere probe is not visible on the reproduced picture, the two signals seen on the normal chromosome 8 (large solid arrowhead) are the sister chromatids, two signals are seen on the aberrant chromosome (open arrowhead).
FISH analysis was kindly provided by S. Jurenka.

megasatellite is present on chromosome 4p, and on the normal chromosome 8, subtelomeric, in approximately band 8p23. Only a single signal, in addition to the centromere specific signal, can be seen on the normal chromosome. The significance of seeing a single signal with FISH analysis, when evidence suggests more than one discrete location of the megasatellite sequences, will be addressed in the discussion.

GS appears to be a classic inv dup(8p) patient. However, presence of a single copy region has not been confirmed. When the megasatellite sequence is used as a FISH probe to the chromosomes of patient GS, there is a maintenance of location and signal intensity on the inversion duplication chromosome when compared to the normal chromosome (figure 32a). TP appears to be a classic inv dup(8p) patient with the exception that at D8S503 and D8S265, which are normally present in single copy on the aberrant chromosome, two maternal alleles were inherited. When the same FISH probe is used against the chromosomes of patient TP there is a more diffuse signal on the aberrant chromosome when compared to the normal chromosome (figure 32b). When the FISH probe is used against the chromosomes of patient MGP, who carries an 8p duplication of unknown orientation, there are two discrete signals on the aberrant chromosome, rather than the single signal seen on the normal chromosome (figure 32c).

Chapter 4: Discussion

The aim of this thesis was to gather molecular evidence supporting the hypothesis of a mechanism of formation of inversion duplication 8p chromosomes mediated by inverted repetitive elements. This mechanism predicts the location of inverted repetitive elements at the cytogenetic center of symmetry of inv dup(8p), located within 8p23.1 in the majority of patients (table 1). Should these repetitive elements be sufficiently distant from each other on the chromosome, a single copy region at the cytogenetic center of symmetry is also predicted. This single copy region was shown to exist by genotyping of 5/5 informative inv dup(8p) patients (Floridia *et al.*, 1996). These data predicted the localization of the proposed inverted repetitive elements at the boundaries of the single copy region, distally in the region between D8S265 and D8S349 and proximally in the region between D8S552 and D8S265 (see figure 3) (Floridia *et al.*, 1996).

4.1 Refinement of the Location of the Distal Repetitive Element Predicted by the Hypothesis

STS genotyping was carried out at marker D8S503, located by YAC content to the region distal to D8S265 but proximal to D8S349 (figure 3), for nine of the inv dup(8p) patients previously reported by Floridia *et al.* (1996). Six of these patients are informative for inheritance of a single D8S503 allele on the inversion duplication chromosome. Previously, a single copy region at the center of symmetry of the inversion duplication chromosome was shown in three of these patients (Floridia *et al.*, 1996). Therefore, the single copy region has

been shown to exist in a total of 8/8 informative patients (Floridia *et al.*, 1996, and table 3). Analysis with D8S503 allowed the location of the distal repetitive element to be refined to the region flanked by D8S503 and D8S349.

4.2 Refinement of the YAC Contig Map in the Region Predicted to Contain the Distal Repetitive Element

A gap between YAC contigs WC8.0 and WC8.1 (MIT/Whitehead Institute) occurs at the region predicted to contain the distal repetitive element. Thirteen YACs, flanking the gap between contigs WC8.0 and WC8.1, were analysed for STS content in order to confirm the MIT/Whitehead Institute data and to refine the map in this region. With the exception that YAC 966b5 was not found to contain marker D8S516, all STS content reported by the MIT/Whitehead Institute was confirmed in our subset of YACs. Additional STS content mapping carried out with markers D8S252, D8S574, and D8S351, allowed restructuring of STS marker order in WC8.1, as well as extension of WC8.1 into the gap between WC8.1 and WC8.0 (figure 9). However, this gap has not yet been spanned.

Refinement of the STS marker order in WC8.1 places D8S503 proximal, but close, to D8S252, on the edge of the gap between contigs (figure 9). D8S252 is not highly polymorphic (heterozygosity = 0.27). Thus, D8S503 is the most distal highly polymorphic marker (heterozygosity = 0.74) in the region spanned by WC8.1, which encompasses the single copy region found in inv dup(8p) patients. Therefore, these data suggest that, of the available polymorphic STS markers, D8S503 is the most useful for determining the distal extent of the single copy region in inv dup(8p) patients.

Although clones spanning the gap between contigs WC8.0 and WC8.1 were not isolated, the STS information collected will be valuable in closing this gap using other clone sources.

Efforts were then focused on analysis of a candidate repetitive element.

4.3 A Candidate Novel Repetitive Element

A novel repetitive element, a megasatellite (MS), was isolated from chromosome 4 clones, and mapped to 4p15 and 8p23 (Gondo *et al.*, 1996, Kogi *et al.*, 1997). The low copy number and localization to 8p23 made it a likely candidate for the proposed repeat, predicted to lie in at least two locations on 8p23. Cosmid, BAC, and YAC clones from chromosome 8, containing the megasatellite sequence, were isolated, mapped, and the organization examined.

4.3.1 Molecular Evidence for the Existence of at Least 4 Megasatellite Locations on 8p

Three different types of megasatellite, contained in cosmid clones, were isolated from the LA08NC01 chromosome 8 cosmid library (figure 15) and analyzed. This library was created from a cell line containing a single chromosome 8 (Wood et al., 1992). The presence of three distinct types of megasatellite in clones isolated from this library can only be explained by the existence of at least three discrete locations of these megasatellites on chromosome 8.

Isolation of BAC and YAC clones containing the megasatellite sequences lent further evidence for at least 3 locations of the megasatellite on chromosome 8 and enabled the placement of the megasatellites on the YAC contig map (figure 19). YAC 967c11 contains STS markers D8S439, D8S227, D8S349 (D8S1819), D8S1935 (figure 19) and a type II megasatellite (table 6, figure 23). YAC 764c7 is a chromosome 8 chimera, containing STS markers D8S439, D8S277, D8S349 (D8S1819), D8S1640 (figure 19), and a type III megasatellite (table 6, figure 23). YAC

773g4 contains STS markers D8S1695, D8S1640, D8S1993 (figure 19), and a type II and type III megasatellite (table 6, figure 23). YAC 799b1 contains STS markers D8S1640, D8S1993, D8S1619, D8S552 (figure 19) and does not contain a megasatellite (figure 23). BAC 87B23 contains STS markers D8S1640, D8S1993, D8S1640 (figure 19), and a type I megasatellite. These data, combined with the knowledge that 3 different megasatellite types exist in a cosmid library constructed from a single chromosome 8, place the megasatellite in at least four discrete locations on chromosome 8.

4.3.2 The Megasatellite is Embedded within a Large Reiterated Sequence

Cross-hybridization between the sequences flanking the different megasatellites was shown by isolation of flanking DNA from type I and type II cosmids followed by hybridization to all megasatellite containing cosmids (table 7, figure 15). Therefore, each class of megasatellite must be embedded in a large reiterated sequence (LRS) and as such, the proposed 'repeat' includes the megasatellite and flanking sequences.

In order to determine the size of the LRS, the insert and ends of BAC 87B23 were hybridized to YACs 967c11, 764c7, 773g4, 799b1. All four YACs share homology with the insert of 87B23, but only 764c7 contains the SP6 end and only 773g4 contains the T7 end (figures 24, 25). YACs 967c11 and 773g4, which map to two discrete regions of the genome, have similar hybridization patterns when the insert of 87B23 is used as a hybridization probe. However, YAC 967c11 failed to hybridize with either end of 87B23. These data can be interpreted in at least two ways: 1. that the LRS is completely contained within BAC 87B23, or 2. it is not completely contained within BAC 87B23 due to a variety of possibilities.

To distinguish between the two possibilities, clones homologous to the ends of 87B23 were isolated (figures 27). Three cosmids, known to contain different types of megasatellites, hybridized with the T7 end of 87B23 (figures 15, 27). Therefore, the LRS is not entirely contained within BAC 87B23. Two BACs, 286D2 and 67B6, extend ~80 kb and ~20 kb, respectively, in either direction from 87B23. However, the partial insert of 286D2 did not hybridize to any of the YACs. And, upon comparison to the hybridization pattern obtained with the insert of 87B23, the insert of 67B6 did not hybridize to any novel *Sst*I fragments in the YACs. One interpretation of these data is, that one endpoint of the LRS is located within 87B23 very near the SP6 end, and, the other endpoint of the LRS is located just beyond the T7 end of 87B23. From these data, it can be concluded that the LRS is at least 165 kb.

A third BAC, 223B23, isolated by library screening, is apparently homologous to 87B23, although the megasatellite is absent (figure 27), and therefore, the three markers D8S1640, D8S1993, and D8S1619, are included in the LRS. The inclusion of these markers in LRS offers an explanation for the hybridization of the PCR product of D8S1993 to two *Eco*RI fragments of YAC 773g4, and for the hybridization of D8S1640 to two *Hind*III fragments of YAC 764c7 (figure 23). These YAC clones may contain two copies of these sequences. Alternatively, these clones may contain copies of these markers that differ from the sequenced PCR products and contain restriction enzyme sites such that digestion separates the sequence onto two fragments.

4.3.4 Orientation of the Megasatellite in BAC 87B23

The orientation of the megasatellite within BAC 87B23 was determined by localization of the T7 end of BAC 87B23, in cosmids that contain the megasatellite and 3′ flanking sequences. BAC 87B23 contains the STS markers D8S1640, D8S1993, and D8S1619, which are ordered

relative to the megasatellite based on YAC content data. Cosmids containing the T7 end of BAC 87B23 do not contain the 5' flanking region of the megasatellite, or the three STS markers. Therefore the orientation of the megasatellite within BAC 87B23, relative to the three markers, can be determined. The orientation of the type I megasatellite on chromosome 8 relative to figure 19 is:

telomere - T7 end - 3′←5′MS - D8S1640 - D8S1993 - D8S1619 - SP6 end - centromere

4.4 The Organization of the Megasatellite on 8p

There are at least 4 locations of the megasatellite on chromosome 8p. However, the organization of the megasatellites on the chromosome is unknown. It is likely that all 8p megasatellites (MS) are embedded within a large reiterated sequence that includes D8S1640, D8S1993, and D8S1619 (which will be referred to as the marker cluster when all three are present in a clone). It is also likely that the YACs 967c11, 764c7 and 773g4 are not chimeric. These YACs contain marker NIB1550, which is located on chromosome 17. YACs 967c11 and 764c7 contain the chromosome 17 marker WI-3139. If indeed all of these YACs are chimeric, it is remarkable and improbable that they would all be chimeric with the same region of chromosome 17. Thus, the inferred chimerism is likely to be an artefact due to amplification within the chromosome 8 LRS corresponding to a paralogous region on chromosome 17. Further, YACs may be unstable, depending on the DNA insert. The inability to isolate clones spanning the gap between the MIT/Whitehead Institute YAC contigs WC8.0 and WC8.1 suggests that this region is unstable in YACs. Therefore, it seems likely that YAC 764c7 is not chimeric for chromosome 8 sequences, but rather that a large deletion occurring in this clone has deleted D8S1935 and the type II megasatellite (figure 33), and that the type III megasatellite in

this YAC clone is actually located in the YAC contig gap between WC8.0 and WC8.1. Therefore, the type III megasatellite would be located in the distal region as well as the proximal region as shown by the presence of type II and III megasatellites on YAC 773g4.

BAC 87B23 contains the marker cluster and a type I megasatellite. The marker cluster is placed on the YAC contig map (figure 33) distal to D8S552. A type I megasatellite was not isolated in YACs. YACs containing megasatellites may be unstable and undergo rearrangements that delete the type I megasatellite, and therefore, the type I megasatellite may be located in the proximal region. Alternatively, the type I megasatellite may be located within the gap between the YAC contigs in the distal region. Both of these explanations are supported by evidence that when the insert of this BAC is used as a hybridization probe, the *Sst*I hybridization pattern differs from the *Sst*I hybridization patterns of YACs 967c11 and 773g4 (figure 24). This suggests that BAC 87B23 may be from a different region of the chromosome than contained within these YACs. Analysis of YAC content of the megasatellite and the marker cluster orients the three markers relative to the megasatellite: MS - 1640 - 1993 - 1619 (figure 19).

4.4.1 The Distal 8p Map Location of the Megasatellite

In the distal region between D8S349 and D8S503 the hypothesis predicts the location of a repetitive element involved in the formation of inv dup(8p) chromosomes. YAC 967c11 contains D8S439, D8S277, D8S349, D8S1935, and a type II MS. It does not contain any of the STS markers from the marker cluster. The presence of D8S349 places this YAC on the genetic map. YAC 764c7 also contains D8S439, D8S277, and D8S349, but does not contain D8S1935. It does, however, contain a type III MS and the STS marker D8S1640. The content of D8S349 places this YAC on the genetic map. Assuming, as discussed above, that this YAC has

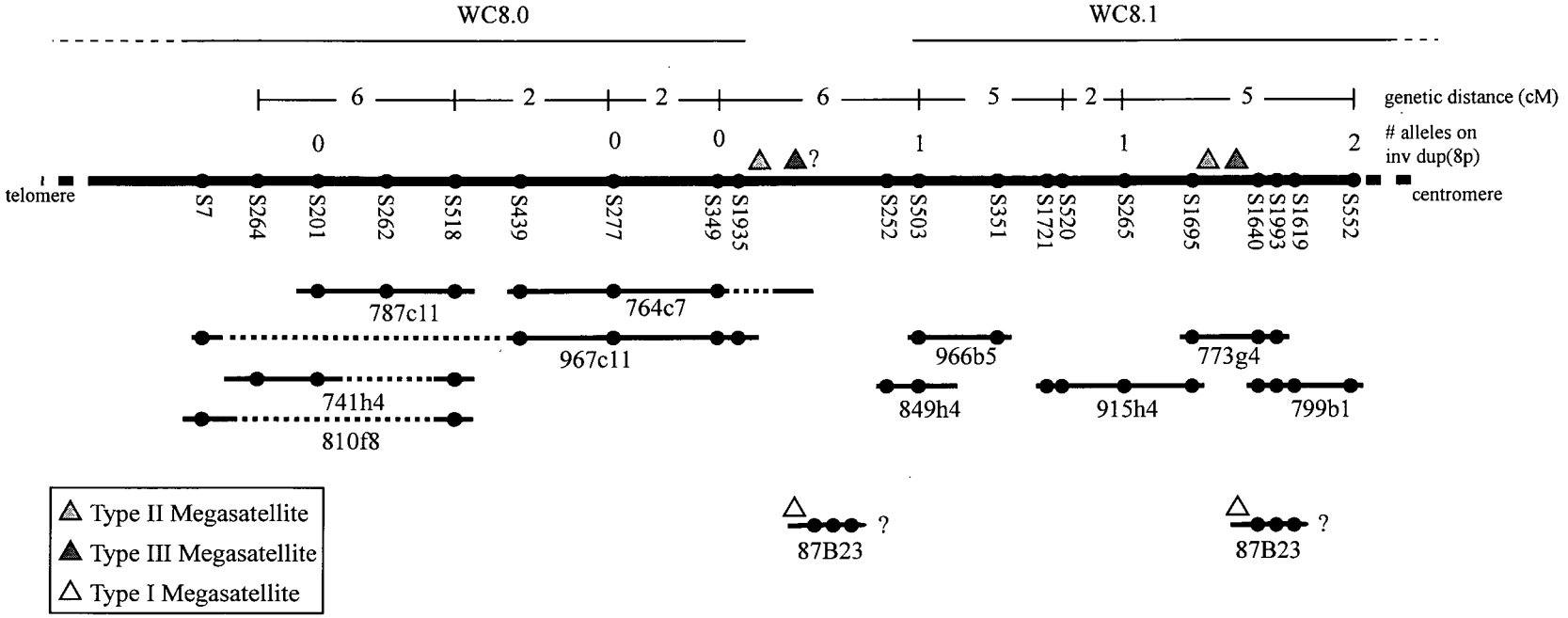


Figure 33: Most likely locations of megasatellite based on YAC/BAC content analysis. Data collected by the MIT/Whitehead Institute (WH) and in our laboratory. WC8.0 and WC8.1 are YAC contigs assembled at the MIT/Whitehead Institute, modifications to STS marker order have been made, additional STS markers have been tested. Dotted lines represent gaps in YACs. All clones, with the exception of BAC 87B23, are YACs. Orientation of megasatellites on chromosome 8 is not known. Each megasatellite is embedded within an large reiterated sequence that includes the STS markers D8S1640, D8S1993, D8S1619. These markers have been placed by YAC content analysis just distal to D8S552, as shown in the map, and will be associated with each copy of the megasatellite (not shown in the figure because orientation relative to each megasatellite is unknown). Location of type I megasatellite is not known.

undergone deletion, then the type III megasatellite may be located proximal to the type II megasatellite in this region. Therefore, the organization in the distal region (figure 33), where the orientation of the bracketed region is unknown, may be:

tel - 349 - 1935 - MSII - 1640 - 1993 - 1619 - (MSIII - 1640 - 1993 - 1619) - 503 - cen

placing the repeat, comprised of the megasatellite and the LRS, in the distal region predicted to contain repetitive elements involved in the formation of inv dup(8p) chromosomes.

4.4.2 The Proximal 8p Map Location of the Megasatellite

In the region between D8S265 and D8S552 the hypothesis predicts the location of a repetitive element involved in the formation of inv dup(8p) chromosomes. YAC 799b1 contains, the marker cluster, D8S552, but no MS. The presence of D8S552 places this YAC on the genetic map. YAC 773g4 contains D8S1695, D8S1640, D8S1993, a type II and a type III MS. The presence of D8S1695 places this YAC on the physical map overlapping YAC 915h4 which contains D8S265 and D8S1695 (figure 33). From these data , the organization in the proximal region (figure 33), where orientation of the bracketed regions is unknown, may be:

tel - D8S265 - D8S1695 - (1619-1993-1640-MSII)/(MSIII)-1640-1993-1619 - D8S552 - cen

placing the repeat, comprised of the megasatellite and the LRS, in the proximal region predicted to contain repetitive elements involved in the formation of inv dup(8p) chromosomes.

4.4.3 Sequences Flanking the Megasatellite are Not Chromosome 8 Specific

An attempt to isolate BAC clones containing the megasatellite, by screening of the Research Genetics BAC library with the T7 and SP6 ends of BAC 87B23, which contains an internal megasatellite, was unsuccessful due to the overwhelming number of positive clones obtained

(>3,000). Attempts to chromosome walk from the ends of BAC 87B23 led to isolation of clones, such as BAC 223B23, that with the exception of the megasatellite, are apparently homologous to the LRS. These data suggested that these clones may be from elsewhere in the genome.

When used as a hybridization probe for FISH analysis, BAC 223B23 hybridizes to multiple locations throughout the genome, including 8p23. The signal on 8p23 is stronger than those seen elsewhere in the genome, likely due to the localization of at least 4 copies of this sequence to this region. Therefore, the sequences flanking the megasatellite are not unique to chromosome 8, and chromosome walking, within this region, will be difficult. The size of LRS is estimated to be at least 165 kb on chromosome 8. However, it will be difficult to estimate the size as one can not be sure that clones are truly overlapping with the chromosome 8 sequences without sequencing. The hybridization of BAC 223B23 to multiple locations throughout the genome supports the assumption that the inferred chimerism of YACs 967c11, 764c7, and 773g4 with chromosome 17 could be due to cross hybridization of the LRS with a paralogous region of chromosome 17.

4.5 The Evolution of 8p Megasatellite

The novel 4p repetitive element, or megasatellite, reported by Kogi *et al.* (1997) is defined by a 4.7 kb *Eco*RI fragment tandemly repeated 12-90 times on a single chromosome 4. This repetitive element has been localized to chromosome 8p23, which suggests that at some time in the past, and perhaps still, the 4p megasatellite was mobile. The 8p megasatellite is comprised of approximately 3 complete copies of the 4p repetitive element, that are contained within a number of different sizes of *Eco*RI fragments. The 4.7 kb fragment is maintained on chromosome 4, but not on chromosome 8, suggesting that the origin of the sequence is chromosome 4p.

Each class of chromosome 8p megasatellite is embedded within a region of sequence that is itself reiterated throughout the genome. This suggests that the LRS was mobile before the megasatellite intruded, and may still be mobile. The LRS, including the embedded megasatellite, is present in at least four locations on chromosome 8. This suggests that the original intrusion of the megasatellite onto chromosome 8 occurred within the reiterated region, followed by dispersion of the LRS and the embedded megasatellite to an additional three locations on chromosome 8p. This dispersion may have occurred by chromosomal interactions, or by transposition of the mobile region. The conclusion that this type of dispersion occurred is supported by sequence evidence that the chromosome 8 type I and type II megasatellites are more closely related to each other than to the 4p megasatellite (Appendix 1).

4.6 Analysis of Patients by STS Genotyping and FISH of the 8p Megasatellite

Previous studies of inv dup(8p) have shown that in all investigated informative cases, there is a deletion of distal 8p (Dill *et al.*, 1987, Barber *et al.*, 1994, Mitchell *et al.*, 1994, de Die-Smulders *et al.*, 1994, Guo *et al.*, 1994, Floridia *et al.*, 1996), and there is a region of single copy located at the cytogenetic center of symmetry (Floridia *et al.*, 1996). As well, all inv dup(8p) chromosomes arise by a maternal interchromosomal event (Barber *et al.*, 1994, Floridia *et al.*, 1996). Five patients (10-14), suspected of carrying an inv dup(8p) chromosome, were genotyped for STS markers mapping to the regions deleted, present in single copy, and duplicated, on the inv dup(8p) chromosome.

FISH analysis with cosmid 153G8, which contains megasatellite sequences, was carried out on patients 10 (TP), 14 (MGP) (Dhooge *et al.*, 1994) and a previously reported inv dup(8p) patient (GS) (Dill *et al.*, 1987, Henderson et al., 1992). If the proposed repetitive element

mediating the formation of inv dup(8p) chromosomes is the repeat containing the megasatellite, then, one would expect a signal on the aberrant chromosome of equal intensity, size, and location from the centromere, as on the normal chromosome (see schematic, figure 5).

4.6.1 Patient GS

Patient GS was included in the analysis as she is believed to be a classic inversion duplication patient, although parental DNA was unavailable for genotyping. Hybridization of cosmid 153G8 containing the megasatellite sequences, to metaphase chromosomes from patient GS, occurred to chromosomes 4 and to the normal and aberrant chromosomes 8. The signals on the chromosomes 8 appeared to be of equal intensity and size, and were located at what appears to be the same distance from the centromere. This analysis supports the hypothesis that the LRS containing the 8p megasatellite, is the repetitive element involved in the formation of inversion duplication(8p) chromosomes.

4.6.2 Patient 10 (TP)

The inv dup(8p) chromosome of patient 10(TP) is derived from a maternal interchromosomal event which led to deletion of distal sequences but did not result in a region of single copy, at the center of symmetry, detectable with the available STS markers (table 8). This is the first report of an inv dup(8p) patient where a single copy region was not detected at the center of symmetry of this chromosome with the currently available markers.

A modification of the proposed hypothesis of a mechanism mediated by inverted repeats would explain how the aberrant chromosome of patient 10(TP) could arise (figure 34). If there are at least two repetitive elements located at the distal breakpoint that are inverted with respect

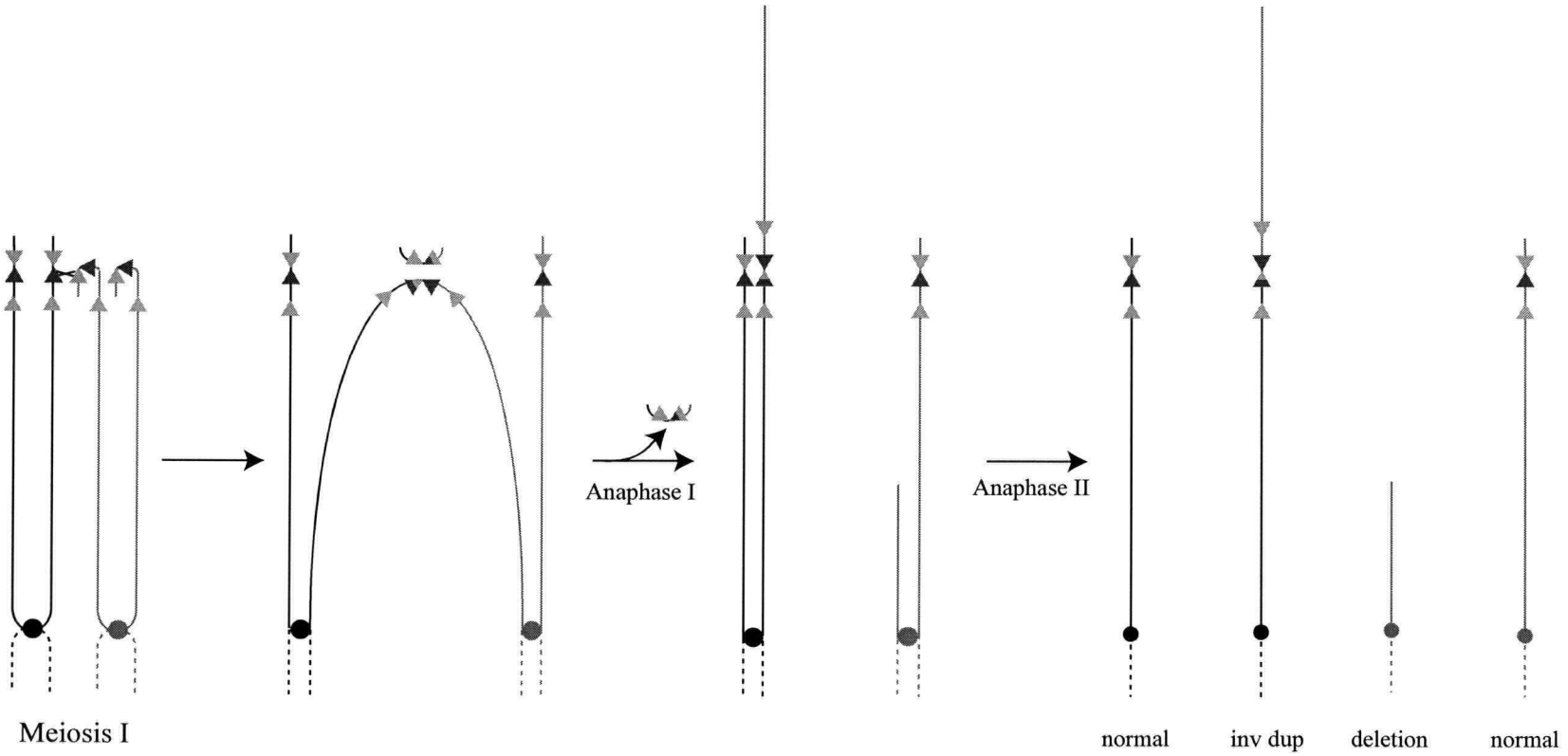


Figure 34: A proposed mechanism for the formation of patient TPs' inversion duplication(8p) chromosome. Illustration is a schematic representation of chromosome interactions. Aberrant pairing at distal inverted repetitive sequences, followed by recombination at the repetitive sequences leads to formation of a dicentric chromosome and acentric fragment. Breakage of the dicentric chromosome, followed by telomere repair results in an inversion duplication chromosome with an additional repetitive sequence, the reciprocal deletion chromosome, and the original normal chromosomes.

to one another, then misalignment of homologues at these sequences, followed by recombination, would result in an inv dup(8p) chromosome with a very small single copy region at the center of symmetry. This single copy region would not be detectable with the currently available STS markers. Further, the repetitive sequences involved in the misalignment would be retained at the same location relative to the centromere, as those on the normal chromosome. However, the repetitive elements in the proximal region would be duplicated, resulting in three regions containing the repetitive sequences on the inv dup(8p) chromosome. Thus, when the repetitive element is used as a FISH probe to patient metaphase chromosomes, one would expect to see a more diffuse signal on the aberrant chromosome, when compared to the normal chromosome.

FISH hybridization of cosmid 153G8 to metaphase chromosomes from patient 10(TP) results in a more diffuse signal on the aberrant chromosome, compared to the normal chromosome. Therefore, FISH hybridization of the cosmid (153G8), containing the megasatellite, supports both the hypothesis that this sequence is involved in the formation of the inv dup(8p) chromosome, and the hypothesis for a modification of the proposed mechanism explained above. (figure 32a). However, it is also possible that other mechanisms, or other repetitive elements, were involved in the formation of this chromosome.

4.6.5 Patient 14 (MGP)

FISH analysis confirms that the extra material on the aberrant chromosome of patient 14 (MGP) is derived from chromosome 8 (Dhooge *et al.*, 1994). In no case were there 3 alleles present at a given marker (table 8). Therefore, there is no genotype evidence for duplication. It is possible that this chromosome originally arose as a result of an intrachromosomal event. To

date, inv dup(8p) chromosomes have only been shown to arise by an interchromosomal event. Karyotype analysis predicts that the duplication encompasses either bands p21.3→p22 or p22→p23.1 (Dhooge *et al.*, 1994). In either case, if the duplication was inverted, FISH of the megasatellite (cosmid 153G8), would result in a single signal on the aberrant chromosome, and if direct, two signals. Two discrete signals are seen on the aberrant chromosome when the megasatellite FISH probe is hybridized to the chromosomes of MGP (figure 32c). Therefore, FISH analysis supports the conclusion that this material is directly duplicated. Classic inv dup(8p) patients are severely mentally retarded and do not reproduce. Patient MGP is not severely mentally retarded, and inherited his aberrant chromosome from his mother, who carried the same chromosome.

Cytogenetic, molecular, and phenotypic differences of this case compared to inv dup(8p) cases supports the conclusion that this patient has a direct, not inverted, duplication of 8p.

4.6.3 Patient 11 (MD)

The inv dup(8p) chromosome of patient 11 (MD) is inherited through a maternal interchromosomal event that led to a deletion of distal sequences. The aberrant chromosome is uninformative for the markers located to the single copy region, and has a duplication of D8S552 (table 8). Genotyping, at the markers tested, suggests that she has a classic inv dup(8p) chromosome. However, the aberrant chromosome is capped by satellites. The mechanism by which this chromosome arose is unknown. Due to the similarity to other inv dup(8p) patients we propose that the underlying mechanism is identical, with the exception that the satellites of one of the acrocentric chromosomes were used to stabilize the aberrant chromosome.

4.6.4 Patients 12 (DC) and 13 (SW)

The aberrant chromosomes of patients 12 (DC) and 13 (SW) are unusual in that the center of symmetry is more distal than that seen in the majority of inv dup(8p) patients (table 1), and a deletion of distal sequences was not detected with the STS markers tested (table 8). The origin of the aberrant chromosome, of both patients, could not be determined from the STS markers tested.

FISH analysis carried out by S. Jurenka, confirms that the extra material on both DC and SW's chromosomes is of chromosome 8 derivation, however, the mechanism by which these inv dup(8p) patients arose is unknown. FISH analysis of normal chromosome 8 with the megasatellite probe (figure 32), and the flanking reiterated sequence probe (figure 30), rules out a location of these sequences distal to 8p23.1. Therefore, the region of duplication containing the megasatellite sequences is unlikely to be involved in the formation of these unusual chromosomes. However, this does not exclude the existence of other repetitive sequences, located in 8p23.2/8p23.3, which could mediate the formation of these chromosomes. Conversely, these chromosomes may arise by a mechanism different than that leading to inv dup(8p) chromosomes with a center of symmetry in 8p23.1.

4.7 FISH as a Tool for Determining the Presence of a Single Copy Region

In 1996, Floridia *et al.* reported evidence for the existence of a single copy region, flanked by the duplication, in inv dup(8p) patients. In 5/16 patients, genotyping was informative for the presence of a single copy region. In an additional 9/16 patients, genotyping was uninformative for copy number. Therefore, in these nine patients, the presence of a single copy region was

demonstrated by FISH of STS PCR products to patient metaphase chromosomes, followed by quantitative analysis of the number of signals present on the aberrant chromosome.

At least four chromosome 8p megasatellites have been located to the distal and proximal regions predicted by the analysis of 8p inversion duplications. When the megasatellite is used as a FISH probe to patient chromosomes, only a single signal is visible on both the normal and aberrant chromosome suggesting that these sequences are within approximately 2-3 Mb of each other and therefore not resolvable as discrete signals. From this, it is reasonable to assume that, depending on the organization at the center of symmetry, duplicated sequences from this region may not appear as discrete signals when used as FISH probes. Therefore, FISH against metaphase chromosomes, without quantification of signal intensity, is not an accurate tool for assessing copy number of STS markers localized to the region between the distal and proximal megasatellites.

4.8 A Single Copy Region May Not be Found in Rare Cases of Inv Dup(8p)

In 1996, Floridia *et al.*, demonstrated the existence of a single copy region located at the center of symmetry of inversion duplication (8p) chromosomes. Currently, this single copy region has been reported in 9/9 informative patients (Floridia *et al.*, 1996, Barber *et al.*, 1998, and table 3). However, Patient 10 (TP), reported in this thesis, has an inversion duplication of bands (p11.2→p23.1) but does not have a region of single copy detectable with the currently available STS markers. D8S265 and D8S503, which lie within the single copy region, show 3 alleles in this patient indicating duplication of the region on the aberrant chromosome. The proposed mechanism could still account for this patient if there are two of the proposed repetitive elements located in close proximity and in opposite orientation near the distal boundary of the

single copy region (see section 4.7.2). In other words, this mechanism would apply if there are multiple copies of the proposed repeat clustering in the regions predicted to contain these repeats. A similar situation may exist on chromosome 15q in the PWS/AS region (Robinson WP, personal communication, Robinson *et al.*, 1997). A gap in the YAC contigs of 8p exists at the region where the additional repeat would lie, therefore, it has not been possible to determine the number of megasatellites in this region, although a re-interpretation of the organization of YAC 764c7 leads to the conclusion that the distal region contains two megasatellites.

Interestingly, cytogenetic analysis suggests that the aberrant chromosome found in patient TP does not differ from other inversion duplication chromosomes with a center of symmetry at 8p23.1. Therefore, the size of this region is less than the size detectable by analysis of metaphase chromosomes, approximately 2-3 Mb.

4.9 Genomic Organization of 8p23.1

FISH analysis of the megasatellite by hybridization of a cosmid containing megasatellite to metaphase chromosomes, resulted in a single signal on the normal chromosome 8 (figure 32). Therefore, the distance between these sequences is below the level resolvable with FISH, estimated to be approximately 2-3 megabases (Mb) (Trask, 1991). YAC analysis places the megasatellite in at least 2 discrete regions of chromosome 8p, located at minimum 8 cM (D8S503-D8S265) and at maximum 18 cM (D8S349-D8S552) apart, based on the sex-averaged Genethon genetic map (figures 3, 19) (Dib *et al.*, 1996, Gyapay *et al.*, 1994). On average, 1 cM is considered to be equal to 1 Mb, however, increased recombination can inflate the genetic distance relative to the physical distance. If 1 cM equals 1 Mb in this region, one would expect these sequences to be located between 8-18 Mb apart. However, evidence from FISH analysis

does not support a region of this size. Therefore, the genetic distance must be inflated in this region. The most obvious explanation for this inflation is that this region is prone to recombination. Examination of the male and female genetic maps suggests that recombination in females, and not males, is increased in this region. The estimate of the size of this region from the female genetic map is at minimum 9.9 cM and at maximum 21.7 cM. Whereas the estimate from the male genetic map is at minimum 2.3 and at maximum 8.7 cM, the lower end of which is consistent with FISH analysis (figure 3). A sex-specific difference in recombination rate in this region may explain why inv dup(8p) chromosomes are of maternal origin. Perhaps some difference in female meiosis allows recombination to occur in this region more frequently than males, and therefore, should these sequences be misaligned during meiosis, recombination would be more likely to occur at these misaligned sequences in females.

A modification of the proposed mechanism mediated by inverted repeats may offer a more robust explanation of how inv dup(8p) chromosomes arise. Intrachromosomal misalignment and recombination at the proposed repetitive sequences, would lead to inversion of the region between the repeats (figure 1). As discussed in the introduction, there is evidence that this type of event occurs in the factor VIII gene and leads to severe haemophilia A (Lakich *et al.*, 1993) and also occurs in the IDS gene and leads to Hunter syndrome (Bondeson *et al.*, 1995). Since this region is located within a single R band, a paracentric inversion of this region would not be detectable at the cytogenetic level. Should a parent carry this inversion, then, at meiosis, the normal process of events to resolve pairing problems inherent to inversions would occur. Namely, a very small inversion loop would form. In the majority of cases recombination would occur outside the loop. However, in rare instances, recombination may occur within the loop, leading to a dicentric and an acentric product, each containing a region of single copy. Breakage

of the dicentric at anaphase, followed by telomere repair, could lead to the formation of an inversion duplication chromosome. The center of symmetry, including the single copy region, would be consistent in unrelated patients, and the extent of duplication would vary. This mechanism is attractive in that it requires only a single aberrant event followed by normal resolution of this event and does not exclude a mechanism based on interchromosomal misalignment of these repetitive sequences.

Analysis of recurrent rearrangements often begins by analysis of novel junction fragments located at the rearrangement breakpoints. Chromosomes arising by the above mechanism would not have novel junction fragments. Therefore, should this mechanism apply, novel junction fragments will not be detected in the regions flanking the single copy region.

4.10 Conclusions

4.10.1 Investigation of Megasatellite as a Candidate for the Repetitive DNA Involved in the Formation of inv dup(8p)

A hypothesis for the mechanism of formation of inv dup(8p) chromosomes predicts that inverted repeats mediate the formation of these chromosomes and, that these repeats are localized to the boundaries of the single copy region found at the center of symmetry of these chromosomes. The proximal boundary of the single copy region was previously localized to the region flanked by D8S265 and D8S552 (Floridia *et al.*, 1996). In this thesis, genotyping of nine patients with STS polymorphic marker D8S503 refines the distal boundary to the region flanked by D8S503 and D8S349. A repeat sequence composed of a megasatellite embedded within large reiterated sequence, termed an LRS, has been identified and localized on the short arm of

chromosome 8 to two regions, one flanked by D8S349 and D8S503, the other flanked by D8S265 and D8S552. The organization of this repetitive DNA on chromosome 8 is unknown. However, individual cosmids derived from a single chromosome 8, and containing one of three types of megasatellites embedded within the LRS, have been found. As well, YAC content analysis predicts 2 discrete locations of the type II megasatellite. Therefore, there at least 4 discrete locations of this repeat sequence (LRS with an embedded megasatellite) on chromosome 8. The minimum size of the repeat has been estimated at 165 kb. When used as a FISH probe, the megasatellite sequences are seen as a single signal on distal 8p, indicating a maximum separation of approximately 2-3 Mb. The genetic distance between the proximal and distal sequences is estimated at 8-18 cM, generally interpreted as 8-18 Mb. This discrepancy in distance between megasatellites would be explained by an increased amount of recombination between these sequences, and appears to be due to increase in recombination in females.

Two hypotheses have been proposed for the mechanism of the formation of inv dup(8p). These mechanisms predict the presence of DNA duplicated on the normal chromosome 8p and located to the region flanked by D8S349 and D8S503, and the region flanked by D8S265 and D8S552. We propose that the repetitive DNA localized to these regions, in this thesis, is the repetitive DNA involved in the formation of inv dup(8p).

4.11 Future Research

Until now, the existence of duplicated DNA in the proposed regions for the formation of the inv dup(8p) chromosomes has been only a hypothesis. The mapping of the megasatellite and flanking duplicated DNA within the regions proposed to contain important sequences for the formation of inv dup(8p) chromosomes allows the investigation of this region from a different

approach than previously thought possible. We now have a candidate sequence that can be studied in a number of ways to determine if it is the sequence involved. Some experiments that will need to be carried out are:

1. Determination of orientation of repeat relative to placed STS markers;
2. Closure of the gap between WC8·0 and WC8·1 and a search for a megasatellite located within this gap;
3. Determination of exact number of repeats and their organization relative to each other;
4. Examination of inv dup(8p) patients for novel junction fragments at the repeat boundaries;
5. Examination of maternal DNA for the presence of novel junction fragments that may indicate the presence of an inversion of the region between D8S349 and D8S552 or/ FISH against stretched maternal DNA to determine the orientation of sequences within this region.

Bibliography

Abeliovich D, Dagan J, Werner M, Lerer I, Shapira Y, Meiner V: Simultaneous formation of inv dup(15) and dup(15q) in a girl with developmental delay: Origin of the abnormal chromosomes. Eur. J. Hum. Genet. 3: 49-55, 1995

Abruzzo MA, Hassold TJ: Etiology of nondisjunction in humans. Environ. Mol. Mutagen 25 Suppl. 26: 38-47, 1995

Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ: Basic Local Alignment Search Tool. J. Mol. Biol. 215: 403-410, 1990

Amos-Landgraf J, Gottlieb W, Rogan PK, Nicholls RD: Chromosome breakage in Prader-Willi and Angelman Syndrome deletions may involve recombination between a repeat at the proximal and distal breakpoints. Am. J. Hum. Genet. Suppl. 55: A38, 1994

Barber JCK, James RS, Patch C, Temple IK: Protelomeric sequences are deleted in cases of short arm inverted duplication of chromosome 8. Am. J. Med. Genet. 50: 296-299, 1994

Barber JCK, Joyce CA, Collinson MN, Nicholson JC, Willatt LR, Dyson HM, Bateman MS, Green AJ, Yates JRW, Dennis NR: Duplication of 8p23.1: a cytogenetic anomaly with no established clinical significance. J. Med. Genet. 35: 491-196, 1998

Bellanne-Chantelot C, Lacroix B, Ougen P, Billault A, Beaufils S, Bertrand S, Georges I, Gilbert F, Gros I, Lucotte G, Susini L, Codani JJ, Gesnouin P, Pook S, Vaysseix G, Lu-Kuo J, Ried T, Ward D, Chumakov I, Le Paslier D, Barillot E, Cohen D: Mapping the whole human genome by fingerprinting yeast artificial chromosomes. Cell 70: 1059-1068, 1992

Birnboim HC, Doly J: A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. 7: 1513-1523, 1979

Bodrug SE, Ray PN, Gonzalez I:. Schmickel RD, Sylvester JE, Worton RG: Molecular analysis of a constitutional X-autosome translocation in a female with muscular dystrophy. Science 237: 1620-1624, 1987

Bodrug SE, Holden JJA, Ray PN, Worton RG: Molecular analysis of X-autosome translocations in females with Duchenne muscular dystrophy. EMBO J. 10: 3931-3939, 1991

Bondeson ML, Dahl N, Malmgren H, Kleijer WJ, Tönnesen T, Carlberg BM, Pettersson U: Inversion of the IDS gene resulting from recombination with IDS-related sequences is a common cause of the Hunter syndrome. Hum. Mol. Genet. 4(4): 615-621, 1995

Boué A, Groopp A, Boué J: Cytogenetics of pregnancy wastage. Adv. Hum. Genet. 14: 1-7, 1985

Brownstein BH, Silverman GA, Little DA, Burke DT, Korsmeyer SJ, Schlessinger D, Plson MV: Isolation of single-copy human genes from a library of yeast artificial chromosome clones. Science 244: 1348-1351, 1989

Buiting K, Greger V, Brownstein BH, Mohr RM, Voiculescu I, Winterpacht A, Zabel B, Horsthemke B: A putative gene family in 15q11-13 and 16p11.2: Possible implications for Prader-Willi and Angelman syndromes. Proc. Nat. Acad. Sci. USA: 89: 5457-5461, 1992

Chance PF, Abbas N, Lensch NW, Pentao L, Roa BB, Patel PL, Lupski JR: Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. Hum. Mol. Genet. 3: 223-228, 1994

Cheng SD, Spinner NB, Zackai EH, Knoll JHM: Cytogenetic and molecular characterization of inverted duplicated chromosomes 15 from 11 patients. Am. J. Hum. Genet. 55: 753-759, 1994

Cheung SW, Sun L, Featherstone T: Molecular cytogenetic evidence to characterize breakpoint regions in Robertsonian translocations. Cytogenet. Cell Genet. 54: 97-102, 1990

Choo KH, Vissel B, Brown R, Filby RG, Earle E: Homologous alpha satellite sequences on human acrocentric chromosomes with selectivity for chromosomes 13, 14, and 21: implications for recombination between nonhomologues and Robertsonian translocations. Nucleic Acids Res. 16: 1273-1284, 1988

Choo, KH, Earle E, McQuillan C: A homologous subfamily of satellite III DNA on human chromosomes 14 and 22. Nucleic Acids Res. 18: 5641-5648, 1990

Chumakov IM, Le Gall I, Billault A, Ougen P, Soularue P, Guillou S, Rigault P, Bui H, De Tand MF, Bariloot E, Abderrahim H, Cherif D, Berger R, Le Paslier D, Cohen D: Isolation of Chromosome 21-specific yeast artificial chromosomes from a total human genome library. Nat. Genet. 1: 222-225, 1992

Clayton-Smith J, Webb T, Cheng XJ, Pembrey ME, Malcolm SE: Duplication of chromosome 15 in the region 15q11-13 in a patient with developmental delay and ataxia with similarities to Angelman syndrome. J. Med. Genet. 30: 529-531, 1993

Cole CG, Goodfellow PN, Bobrow M, Bentley DR: Generation of novel sequence tag sites (STSs) from discrete chromosomal regions using *Alu*-PCR: Genomics 10: 816-826, 1991

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

de Die-Smulders CEM, Engelen JJM, Schrander-Stumpel CTRM, Govaerts LCP, de Vries B, Vles JSH, Wagemans A, Schijns-Fleuren S, Gillessen-Kaesbach G, Fryns JP: Inversion duplication of the short arm of chromosome 8: Clinical data on seven patients and review of the literature. Am. J. Med. Genet. 59: 369-374, 1995

Deaven LL, Van Dilla MA, Bartholdi MF, Carrano AV, Cram LS, Fuscoe JC, Gray JW, Hildebrand CE, Moyzis RK, Perlman J: Construction of human chromosome-specific DNA libraries from flow-sorted chromosomes. Cold Spring Harb. Symp. Quant. Biol. 51: 159-167, 1986

Debella L, Schertzer M, and Wood S: Identification of a Novel Human Gene (GOR) localized to 8q13-8q22. unpublished, 1998

Dib C, Faure S, Fizames C, Samson D, Drouot n, vignal A, millasseau P, Marc s, Hazan j, Seboun E, Lathrop m, Gyapay G, Morisette J, Weissenbach J: A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380: 152-154, 1996

Dill FJ, Schertzer M, Sandercock J, Tischler B, Wood S: Inverted tandem duplication generates a duplication deficiency of chromosome 8p. Clin. Genet. 32: 109-113, 1987

Dhooge C, Van Roy N, Crean M, Speleman F: Direct transmission of a tandem duplication in the short arm of chromosome 8. Clin. Genet. 45: 36-39, 1994

Earle E, Shaffer LG, Kalitsis P, McQuillan C, Dale S, Choo KHA: Identification of DNA sequences flanking the breakpoint of human t(14q21q) Robertsonian translocations. Am. J. Hum. Genet. 50: 717-724, 1992

Engelen JJM, de Die-Smulders CEM, Fryns JP, Hoovers JMN, Albrechts JCM, Loots WJG, Jacobs ME, Hamers AJH: Partial trisomy and monosomy 8p due to inversion duplication. Clin. Genet. 45: 203-207, 1994

Evans GA, Lewis KA, Rothenberg BE: High efficiency vectors for cosmid microcloning and genomic analysis. Gene 79: 9-20, 1989

Feldman GL, Weiss L, Phelan MC, Schroer RJ, Van Dyke DL: Inverted duplication of 8p: ten new patients and review of the literature. Am. J. Med. Genet. 47: 482-486, 1993

Feinberg AP, Vogelstein B: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132: 6-13, 1984a

Feinberg AP, Vogelstien B: Addendum: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 137: 266-267, 1984b

Ferguson-Smith MA: Chromosome satellite association. Lancet i: 1156-1157, 1967

Flint J, Craddock CF, Villegas A, Bentley DP. Williams HJ, Galanello R, Cao A, Wood WG, Ayyub H, Higgs DR: Healing of broken human chromosomes by the addition of telomeric repeats. Am. J. Hum. Genet. 55: 505-512, 1994

Floridia, G, Piantanida, M, Minelli A, Dellavecchia C, Bonaglia C, Rossi E, Gimelli G, Croci G, Franchi F, Gilgenkrantz S, Grammatico P, Dalprà L, Wood S, Danesino C, Zuffardi O: The same molecular mechanism at the maternal meiosis I produces mono- and di-centric 8p duplications. Am. J. Hum. Genet. 58: 785-796, 1996

Fryns JP, Kleczkowska A, Dereymaker AM, Hoefnagels M, Geremans G, Marien J, Van den Berghe H: Partial 8p trisomy due to interstitial duplication: karyotype: 46,XX, inv dup(8) (p21.1→p22). Clin. Genet. 28: 546-549, 1985

Fryns JP, Kleczkowska A, Vogels A, van den Berghe H: Normal phenotype and slight mental retardation in de novo distal 8p deletion (8pter-8p23.1). Ann. Genet. (Paris) 32: 171-173, 1989

Fuscoe JC, Clark LM, Van Dilla MA: Construction of fifteen human chromosome-specific DNA libraries from flow-purified chromosomes. Cytogenet. Cell Genet. 43: 79-86, 1986

Giacalone JP, FranckeU: Common sequence motifs at the rearrangement sites of a constitutional X/autosome translocation and associated deletion. Am. J. Hum. Genet. 50(4): 725-741, 1992

Gobin SJ, Alcaide-Loridan C, Bono MR, Ottone C, Chumakov I, Rothstein R, Fellows M: Transfer of yeast artificial chromosome into mammalian cells and comparative study of their integrity. Gene 163(1): 27-33, 1995

Gondo Y, Hando Y, Saitoh T, Okada M, Kogi S, Fukushiga S, Ikeda JE: Megasatellite DNA: A novel 4.7 kb tandem repetitive sequence found in human 4p15. Am. J. Hum. Genet. Suppl. 59(4): A303, 1996

Gorinati M, Caufin D, Minelli A, Memo L, Gaspardo G, Dodero A : Inv dup (8)(p21.1→22.1): further case report and a new hypothesis on the origin of the chromosome abnormality. Clin. Genet. 39: 55-59, 1991

Grunstein M, and Hogness DS: Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Nat. Acad. Sci. 72:3961, 1975

Guichaoua MR, Devictor M, Hartung M, Luciani JM, Stahl A: Random acrocentric bivalent associations in human pachytene spermatocytes: molecular implications in the occurrence of Robertsonian translocation. Cytogent. Cell. Genet. 42: 191-197, 1986

Guo WJ, Callif-Daley F, Zapata MC, Miller ME: Clinical and cytogenetic finding in seven cases of inverted duplication of 8p with evidence of a telomeric deletion using fluorescence in situ hybridization. Am. J. Med. Genet. 58: 230-236, 1995

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, Bernardi G, Lathrop M, Weissenbach J: The 1993-94 Genethon human genetic linkage map. Nat. Genet. 7(2): 246-339, 1994

Han JY, Choo KHA, Shaffer LG: Molecular cytogenetic characterization of 17 rob(13q14q) Robertsonian translocations by FISH, narrowing the region containing the breakpoints. Am. J. Hum. Genet. 55: 960-967

Henderson KG, Dill FJ, Wood S: Characterization of an inversion duplication of the short arm of chromosome 8 by fluorescent in situ hybridization. Am. J. Med. Genet. 44: 615-618, 1992

Hongell K, Knuutila S, Westermarck T: Two cases of an abnormal short arm of chromosome 8 (8p+) associated with mental retardation. Clin. Genet. 13: 237-240, 1978

Hoo JJ, Chao M, Szego K. Rauer M, Echiverri SC, Harris C: Four new cases of inverted terminal duplication: A modified hypothesis of mechanism of origin. Am. J. Med. Genet. 58: 299-304, 1995

Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu S-H, Hu X, Colbert AME, Rosenberg c, Reeve-Daly MP, Rozen S, Hui L, Wu X, Vestergaard C, Wilson km, Bae JS, Maitra S, Ganiatsas S, Evans CA, DeAngelis MM, Ingalls KA, Nahf RW, Horton Jr. LT, Anderson MO, Collymore AJ, Ye W, Kouyoumijian V, Zemsteva IS, Tam J, Devine R, Courtney DF, Renaud MT, Nguyen H, O'Connor TJ, Fizames C, Faure S, Gyapay G, Dib C, Morissette J, Orlin JB, Birren BW, Goodman N, Weissenbach J, Hawkins TL, Foote S, Page DC, Lander ES: An STS-based map of the human genome. Science 270: 1945-1954, 1995

Inoue H, Nojima H, Okayama H: High efficiency transformation of *Escherichia coli* with plasmids. Gene 96: 23-28, 1990

Ish-Horowicz D and Burke JF: Rapid and efficient cosmid cloning. Nuc. Acid Res. 9: 2989, 1981

Jacobs, PA: The role of chromosome abnormalities in reproductive failure. Reprod. Nutr. Dev. Suppl. 1:63S-74S, 1990

Jauch A, Daumer C, Lichter P, Murken J, Schroeder-Kurth T, Cremer T: Chromosomal *in situ* suppression hybridization of human gonosomes and autosomes and its use in clinical cytogenetics. Hum. Genet. 85: 145-150, 1990

Jensen PKA, Junien C, Despoisse S, Bernsen A, Thelle T, Friedrich U, de la Chapelle A: Inverted tandem duplication of the short arm of chromosome 8: A non-random de novo structural aberration in man. Localization of the gene for glutathione reductase in subband 8p21.1. Ann. Genet. 25(4): 207-211, 1982

Kalitsis P, Earle E, Vissel B, Shaffer LG, Choo KHA: A chromosome 13-specific human satellite I DNA subfamily with minor presence on chromosome 21: further studies of Robertsonian translocations. Genomics 16: 104-112, 1993

Kleczkowska A, Fryns JP, D'Hondt F, Jaeken J, van den Berghe H: Partial duplication 8p due to interstitial duplication: inv dup (8)(p21.1→p22). Ann. Genet. 30(1): 47-51, 1987

Kogi M, Fukushigi S, Lefevre C, Hadano S, Ikeda JE: A novel tandem repeat sequences located on human chromosome 4p: isolation and characterization. Genomics 42(2): 278-283, 1997

Lakich D, Kazazian HH Jr., Stylianos EA, Gitschier J: Inversion disrupting the factor VIII gene are a common cause of severe haemophilia A. Nat. Genet. 5: 236-241, 1993

Leana-Cox J, Jenkins L, Palmer C, Plattner R, Sheppard L, Flejter WL, Zackowski J, Tsein F, Schwartz S: Molecular cytogenetic analysis of inv dup(15) chromosomes; using probes specific for the Prader-Willi/Angelman syndrome region: clinical implications. Am. J. Hum. Genet. 54: 748-756, 1994

Lopes J, Ravise N, Vandeberghe A, Palau F, Ionasescu V, Mayer M, Levy N, Wood N, Tachi N, Bouche P, Latour P, Ruberg M, Brice A, LeGuern E: Fine mapping of de novo CMT1A and HNPP rearrangements within CMT1A-REPs evidences two distinct sex-dependent mechanisms and candidate sequences involved in recombination. Hum. Mol. Genet. 7(1): 141-148, 1998

Lupski JR, de Oca-Luna RM, Slaugenhaupt S, Pentao L, Guzzetta V, Trask B, Sucedo-Cardenas O, Barker DF, Killian JM, Garcia CA, Chakravarti A, Patel PI: DNA duplication associated with Charcot-Marie-Tooth disease type 1A. Cell 66: 219-232, 1991

Malmgren H, Carlberg BM, Pettersson U, Bondeson ML: Identification of an alternative transcript from the human iduronate-2-sulfatase (IDS) gene. Genomics 29(1): 291-293, 1995

Mascari MJ, Gottlieb W, Rogan PK, Butler MG, Waller DA< Armour JAL, Jeffrey AJ, Ladda RL, Nicholls RD. The frequency of uniparental disomy in Prader-Willi syndrome. New Engl. J. Med. 326: 1599-1607, 1992

Mattei JF, Mattei MG, Ardissone JP, Coignet J, Giraud F: Clinical, enzyme, and cytogenetic investigations in three new cases of trisomy 8p. Hum. Genet. 53:315-321, 1980

Mears AJ, Duncan AMV, Budarf ML, Emanuel BS, Sellinger B, Siegel-Bartelt J, Greenberg CR, McDermid HE: Molecular characterization of the marker chromosome associated with Cat Eye syndrome. Am. J. Hum. Genet. 55: 134-142, 1994

Minelli A, Floridia G, Rossi E, Clementi M, Tenconi R, Camurri L, Bernardi F, Hoeller H, Previde Re C, Maraschio P, Wood S, Zuffardi O, Danesino C: D8S7 is consistently deleted in inverted duplications of the short arm of chromosome 8 (inv dup 8p). Hum. Genet. 92: 391-396 1993

Mitchell JJ, Vekemans M, Luscombe S, Hayden M, Weber B, Richter A, Sparkes R, Kojis T, Watters G, Der Kaloustian VM: U-type exchange in a paracentric inversion as a possible mechanism of origin of an inverted tandem duplication of chromosome 8. Am. J. Med. Genet. 49: 384-387, 1994

Morin GB: Recognition of a chromosome truncation site associated with α-thalassaemia by human telomerase. Nature 353: 454-456. 1991

Mutirangura A, Kuwano A, Robinson WP, Greenberg F, Malcolm S, Ledbetter DH. Duplication of chromosome 15q Prader-Willi and Angelman syndromes: a gene dosage paradox. Am. J. Hum. Genet. Suppl. 53: A584, 1993

Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, Ledbetter DH, Caskey CT: *Alu* polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. Proc. Natl. Acad. Sci. USA 86: 6686-6690, 1989

Nevin NC, Morrison PJ, Jones J, MacC Reid M: Inverted tandem duplication of 8p12→p23.1 in a child with increased activity of glutathione reductase. J. Med. Genet. 27: 135-136, 1990

Nicholls RD, Fischel-Ghodsian N, Higgs DR: Recombination at the human α-globin gene cluster: sequence features and topological constraints. Cell 49: 369-378, 1987

North KN, Wu BL, Cao BN, Whiteman DAH, Korf BR: CHARGE association in a child with de novo inverted duplication (14)(q22→q24.3). Am. J. Med. Genet. 57: 610-614, 1995

Page SL, Shin JC, Han JY, Choo KH, Shaffer LG: Breakpoint diversity illustrates distinct mechanisms for Robertsonian translocation formation. Hum. Mol. Genet. 5(9): 1279-1288, 1996

Poloni L, Maraschio P, Danesino, Zuffardi O: Two cases with a de novo inverted duplication of a segment of chromosome 8. Clin. Genet. 19: 530, 1981

Raeymaekers P, Timmerman V, Nelis E, Van Hul W, De Jonghe P, Hoogendijk JE, Baas F, Bakker DF, Martin JJ, De Visser M, Bolhuis PA, Van Broeckhoven C, and the HMSN Collaborative Research Group: Duplication in chromosome 17p11.2 in Charcot-Marie-Tooth neuropathy type 1A (CMT1A). Neuromusc. Disord. 1: 93-97, 1991

Redha AA, Murthy DS, Kandil H, Farag TI, Usha R, al-Awadi SA, Jeryan LA, al-Nagdy K, el-Ghanem M: Partial trisomy of short arm of chromosome 8 (46,XY,inv dup(8)(p21→pter) in a

Bedouin child with multiple congenital anomalies and mental retardation. Indian J. Pediatr. 61(3): 301-306, 1994

Rethoré M, Aurias A, Couturier J, Dutrillaux B, Prieur M, Lejeune J: Chromosome 8: trisomie complète et trisomies segmentaires. Ann. de Génét. 20: 5-11, 1977

Riley J, Butler R, Ogilvie G, Finniear R, Jenner D, Powell S, Anand R, Smith JC, Markham AF: A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. Nuc. Acids Res. 18: 2887-2890, 1990

Robinson WP, Bottani A, Yagang X, Balakrishman J, Binkert F, Mächler M, Prader A, Schinzel A: Molecular, cytogenetic, and clinical investigations of Prader-Willi syndrome patients. Am. J. Hum. Genet. 49: 1219-1234, 1991

Robinson WP, Binker F, Gin R, Vazquez C, Müller W, Rosenkranz W, Schinzel A: Clinical and molecular analysis of five inv dup(15) patients. Eur. J. Hum. Genet. 1: 37-50, 1993

Robinson WP, Horsthemke B, Leonard S, Malcolm S, Morton C, Nichols RD, Ritchie RJ, et al., Report of the third international workshop on human chromosome 15 mapping 1996. Cytogenet. Cell Genet. 76: 1-13, 1997

Robinson WP, Dutly F, Nicholls RD, Bernasconi F, Penaherrera M, Michaelis RC, Abeliovich D, Schinzel AA: The mechanisms involved in formation of deletions and duplications of 15q11-q13. J. Med. Genet. 35(2): 130-136, 1998

Rose MD, Winston F, Hieter P: Methods in yeast genetics - A laboratory course manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1990

Saitoh S, Harada N, Jinno Y, Hashimoto K, Imaizumi K, Kuroki Y, Fukushima Y, Sugimoto T, Renedo M, Wagstaff J, Lalande M, Mutirangura A, Kuwano A, Ledbetter D, Niikawa N.: Molecular and clinical study of 61 Angelman syndrome patients. Am. J. Med. Genet. 52: 158-163, 1994

Sambrook J, Fritsch FF, Maniatis R: Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 1989

Schmickel RD and Knoller M: Characterization and localization of the human genes for ribosomal ribonucleic acid. Pediatr. Res. 11: 929-935, 1977

Schreck RR, Breg WR, Erlanger BF, Miller OJ: Preferential derivation of abnormal human G-group-like chromosomes from chromosome 15. Hum. Genet. 36: 1-12, 1977

Sonnhammer E.L.L. and Durbin R: A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 167:GC1-10, 1995

Southern EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98: 503-517, 1975

Stetten G, Charity LL, Kasch LM, Scott AF, Berman CL, Pressman E, Blakemore KJ: A paternally derived inverted duplication of 7q with evidence of a telomeric deletion. Am. J. Med. Genet. 68(1): 76-81, 1997

Sullivan BA, Jenkins LS, Karson EM, Leana-Cox J, Schwartz S: Evidence for structural heterogeneity from molecular cytogenetic analysis of dicentric Robertsonian translocations. Am. J. Hum. Genet. 59: 167-175, 1996

Taylor KM, Francke U, Brown MG, George DL, Kaufhold ML: Inverted tandem ("mirror") duplications in human chromosomes: inv dup 8p, 4q, 22q. Am. J. Med. Genet. 1: 3-19, 1977

Telvi L, Ion A, Carel JC, Desguerre I, Piraud M, Boutin AM, Feingold J, Ponsot G, Fellous M, McElreavey K: A duplication of distal Xp associated with hypogonadotrophic hypogonadism, hypoplastic external genitalia, mental retardation, and multiple congenital abnormalities. J. Med. Genet. 33(9): 767-771, 1996

Therman E: Human chromosomes: structure, behavior, effects, 1st ed. Springer, New York, Heidelberg, and Berlin, 1980

Therman E, Susman B, Denniston C: The nonrandom participation of human acrocentric chromosomes in Robertsonian translocations. Ann. Hum. Genet. 53:49-65, 1989

Thompson J, Gibson T, Thompson J., Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673-4680, 1994

Tomfohrde J, Wood S, Schertzer M, Wagner MJ, Wells DE, Parrish J, Sadler LA, Blantor SH, Daiger SP, Wang Z: Human chromosome 8 linkage map based on short tandem repeat polymorphisms: effect of genotyping errors. Genomics 14(1): 144-152, 1992

Trask BJ: Fluorescent *in situ* hybridization applications to cytogenetics and gene mapping. T.I.G.S. 7(5): 149-154, 1991

Trowell HE, Nagy A, Vissel B, Choo KHA: Long-range analysis of the centromeric regions of human chromosomes 13, 14, and 21: identification of a narrow domain containing two key centromeric DNA elements. Hum. Mol. Genet. 2: 1639-1649, 1993

Tuck-Muller CM, Martinez JE, Batista DA, Kearns WG, Wertelecki W: Duplication of the short arm of the X chromosome in mother and daughter. Hum. Genet. 91(4): 395-400, 1993

Ullmann A, Jacob F, Monod J: Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the β-galactosidase structural gene of *Escherichia coli*. J. Mol. Biol. 24: 339, 1967

Urbán Z, Helms C, Fekete G, Csiszár K, Bonnet D, Munnich A, Donis-Keller H, Boyd CD: 7q11.23 deletions in Williams syndrome arise as a consequence of unequal meiotic crossover. Am. J. Hum. Genet. 59: 958-962, 1996

Van Dyke DL: Isochromosomes and interstitial tandem direct and inverted duplications. In The Cytogenetics of Mammalian Autosomal Rearrangements. Alan R. Liss, Inc. 635-665, 1988

Verma and Babu: Human chromosomes: manual of basic techniques. Elmsford, NY: Pergamon Press, 1989

Vocero-Akbani A, Helms C, Wang JC, Sanjurjo J, Korte-Sarfaty J, Veile RA, Liu L, Jauch A, Burgess AK, Hing AV, Holt MS, Ramachandra S, Whelan AJ, Anker R, Ahrent LL, Chen M, Gavin MR, Iannantuoni K, Morton SM, Pandit SD, Read CM, Steinbrueck T, Warlick C, Smoller DA, Donis-Keller H: 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, 1996

Wagner MJ, Ge Y, Siciliano M, Wells DE: A hybrid cell mapping panel for regional localization of probes to human chromosome 8. Genomics 10: 114-125, 1991

Wandstrat AE, Leana-Cox J, Jenkins L, Schwartz S: Molecular cytogenetic evidence for a common breakpoint in the largest inverted duplications of chromosome 15. Am. J. Hum. Genet. 62: 925-936, 1998

Warburton D: Outcome of cases of de novo structural rearrangements diagnosed at amniocentesis. Prenatal Diagn. 4: 69-80, 1987

Weleber RG, Verma RS, Kimberling WJ, Fieger HG Jr, Lubs HA: Duplication-deficiency of the short arm of chromosome 8 following artificial insemination. Ann. Genet. 19(4): 241-247, 1976

Wilson OJ, Meaney CA, Hopwood JJ, Morris CP: Sequence of the human iduronate 2-sulfatase (IDS) gene. Genomics, 17: 773-778, 1993

Wolff DJ, Schwartz S: Characterization of Robertsonian translocations by using fluorescence in situ hybridization. Am. J. Hum. Genet. 50: 174-181, 1992

Wood S, Poon R, Ridell DC, Royle NJ, Hamerton JL: A DNA marker for human chromosome 8 that detects alleles of differing sizes. Cytogenet. Cell Genet. 42: 113-118, 1986

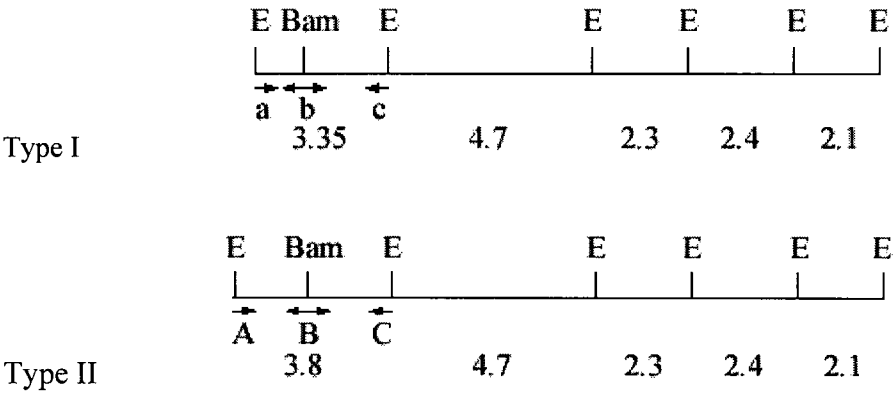
Wood S, Schertzer M, Drabkin H, Patterson D, Longmire JL, Deaven LL: Characterization of a human chromosome 8 cosmid library constructed from flow-sorted chromosomes. Cytogenet. Cell Genet. 59: 243-237, 1992

Wood S, Schertzer M, Yaremko ML: Sequence identity located CEBPD and FGFR1 to mapped human loci within proximal 8p. Cytogenet. Cell Genet. 70: 188-191, 1995

Yen PH, Li XM, Tsai SP, Johnson C, Mohanda T, Shapiro LJ: Frequent deletions of the human X chromosome distal short arm result from recombination between low copy repetitive elements. Cell 61: 603-610, 1990

Zackowski JL, Nicholls RD, Gray BA, Bent-Williams A, Gottlieb W, Harris PJ, Waters MF, Driscoll DJ, Zori RT, Williams CA. Cytogenetic and molecular analysis in Angelman syndrome. Am. J. Med. Genet. 46: 7-11, 1993

Appendix 1: Sequence Comparisons



Location of sequence within the type I and type II 8p megasatellites. E = *Eco*RI, Bam = *Bam*HI, *Eco*RI fragment sizes are in kb. Alignments were made using clustalw, with reference to BLAST alignment searches.

a. Comparison of sequences A, a, and the 4p megasatellite.

```
type II     TNTTATTAAAGAAAAAAATACGCTGTGCTAAATACCATAANTTCATTGACTAATCTCAGG 332
type I      TATTATTAAGGAAGAAAATACGCTGTGCTAAATACTATACTTCCATTGACTATTCTCAGG 324
4p          TCTCACTAGGGAAGAAAATACGCTGTGCTAAATACTATACTT-CATTGACTATTCTCAGG 4423
            * * * **  *** ********************** ***  * ********* *******

type II     TCAGAAAGCACACTTCCGATTTCTTGTCCTTCT-GTCGCTGAGAGGATGATGATAGCTGC 273
type I      TCAGAAAGCGCACTTCAGACTTCTTGTCCTTCCCGTTGATGAGAGGATGACGGTAGCTGC 264
4p          TCAGAAAGCGCACTTTCGACTTCTTGTCCTTCC-GTCGCTGAGAGGATGATGGCAGCTGC 4482
            ********* *****  ** ************  ** * *********** *  ******

type II     CAAAAGTACATA-CTTGGAAGTTCATCCCAGCACGAGCACACACACACATAAACACACAC 214
type I      CAAAAGTACATAACTTGGAAGTTCATCCCAGCACAAGCACACACACACA----------C 214
4p          CAAAAGTACCTA-CTTGGAGGTTCATCCCAGCACAAACACACACACACA--------CAC 4533
            ********* ** ****** ************** * ************          *

type II     ACACACACACACACACACACACACACACACACACACACACAGACACACACAGGGTTTCATAGGT 154
type I      ACACAAACACACACACACACACACACACACAGAGAGAGATACACACACGGTTTCATAGGT 154
4p          GCCCCCCCACACACACACACAAACACACTCACACACACACACGCACACGGTTTCCTAGGT 4593
             * *   ************** ****** ** * * * * ** **** ****** *****

type II     AAAGATTTCTTCCCTGACATTCTTTTACCTAAAATAAGGCAACTGTGCGGCCACTGCCCA 94
type I      AAAGATTTCTTCCCTGACATTCTTTTACCTAAAATAAGGCAACTGTGTGGCCACTGTCCC 94
4p          AAAGATTTCTTCCCTGCCATTGCTTTACCTAAAATAAGGCAACTGTGAGGCCGCTGTCCC 4653
            ***************** ****  ************************ **** *** **
```

```
type II     AACCCGGTTACACTCATATTATATGTGCCTATCACCCTGAGGAGTAATTTGATTCAGGTG 34
type I      AACCCGGTTACACTCATATTATATGTGCCTATCACCCTGAGGAGTAATTTGATTCAGGTG 34
4p          AACCCGGTTACACTCCTATTATATGTGCCTATCATCCTGAGGAGTAATTTGATTCAGGTG 4713
            *************** ****************** ************************

type II     TTCTAGAAGTCATGATGTGGGCTGTGTCTGTTG     1
type I      TTCTAGAAGTCATGATGTGGGCTGTGTCTGTTG     1
4p          TTCTGGAAGTCATGCTGTGGGCTGTGTCTGTTG     4746
            **** ********* ******************
```

b. Comparison of sequences B, b, and the 4p megasatellite. The *Bam*HI site is shown in bold.

Alignment b1:

```
type II     -AATTNTCTGAGGAATGCAAGAGGATACAACCTAAGACAAAAAACTTAATTGAATCCTGA 406
type I      -AATTNTCTGAGGAATGCAAGAGGATACAACCTAAGACAAAAAACTTAATTGAATCCTGA 405
4p          -AATTCCCAGCG--ATGCCAGGGGACACACCCTGTGACTCCTTCCTGAATTGAGTGCTGA 57
             ****  * * *  **** ** *** *** ***  ***      ** ****** * ****

type II     TATTTCATTAGTAAATAGGGTAATTGATGGATAAATGTAATGGTCTCGGTGGGTGGACAG 466
type I      TATATTATTAGTAAATAGGGTAATTGATGGATAAATGTAATGGTCTCGGTGGGTGGACAG 465
4p          TATTTGATTGGCTTATCGCGCACCTGATGAGTGGGTGGGGGTGTTCGCGGTTGGTGGGGGT 117
            *** * *** *   ** * * *  *****  *   **   ** ** **** *****

type II     TAGTTAAATAAGGGCTGATGCAGCAAGATAATTATTTAAAGGCGTTTGAAAGAAA---TT 521
type I      TAGTTATATAAGGGCTGATGCAGCAAGATAATTATTTAAAAGAGTTTGAAAGAAA---TG 522
4p          GACTTACAGAAGGGCTGATGCGGCCAGAGAGCTCGTCA------TTTGAA-GACTCTCTC 170
             * *** * ************ ** *** *  *  * *    ****** **     *

type II     GAAACAGGAGAGTGGATGTATTCAGCTAAAATAAAATCCGGAAGCCCTGAA-ATAAATNT 582
type I      GAAACAGAAGAGTGGATGTATTCAGCTAAAATAAAATCCGGAAGCCCTGAA-ATAAATNT 581
4p          GGAAG-GGATAGCGTCT---TTCTGC----A--ACCTGCGG--TCCCAGCAGACAAACCT 218
            * **  * * ** *  *   *** **    *  *  * ***   *** * * * ***  *

type II     CATTTTGCGGGTAAAAAAATGGCATTGGAGGA-GATTCTGGGTCAATCATG-AAGCTGTG 640
type I      CATTTTGGGTGTAAAAAAATGGCATTAGAGGA-GATTCTGGGTCAATCATCCAAGCTGTG 640
4p          TGTGATCCTCGTTCCA--GTCGACATGGAGGACGACTCAC--TCTA-CTTGGGAGGTG-- 271
              *  *    **   *   * *   * ***** ** **    ** * * *   ** **

type II     AAAGTTGCATCTTGGAAGCA GGATCC CTGTAATGAAACGAGACTTGTTTATCAGAGGTGG 700
type I      AAAGCTGCATCTTGGAAGCA GGATCC CTGTAATGCAACGAGACTTGTATATCAGAGGTGG 700
4p          --AGTGGCAGTT--CAACCACTTTCC----AA---AAC-----------TCACATCTTC 308
              **  ***  *   ** **   ***    **   ***            *** *  *

type II     TCTTTCAGANGAAAANATTTNGAAGAATGGCCCCTTCCTTTTGTGTATTTGACGATTAGA 760
type I      TCTTTCAGTGGAAAAGATTTTGAAGAATGGACCCTTCCTTTTGTGTATTTGACGATTAGA 760
4p          TCGGCCCGATGCA--GCTTTTGCTGAAATCCAGCGGACTTCTCT--------CCCTGAGA 358
            **   * *  * *    *** *  ***      *   *** * *        *  * ***
```

```
type II     TNTCATGCCAAATTTCGGGTTTTAAACTCTATTTAANCNTTANAA-NAATTANCTANAAT 819
type I      CTTCATGCCAAATCTCGGGTTTAAAACTCTATTTAAACATTAACA-GAATTAATTAAAAT 819
4p          AGTCA---CCACTCTCATGT--GAGACCCGTGTCGACCTCTGTGACGATTTGGCTCCTGT 413
              ***   * * * **  **   * ** *   *  * * *   *  * **   *    *

type II     GGCCAAAAAACAAGAAATTTTTTGATT--AGGAATCGTCAAATATTCATTTCTTGTTAGA 877
type I      GGCAAAAAATCAAGAAATTTTTTGATT--AGGAATCATCACATATTCATTTCTTGTTAGG 877
4p          GGCAAGACAGC----------TTGCTCCCAGGGAGAAGC----------TTCCTCTGAG- 452
            *** * * * *          *** *   *** *    *          *** * * **

type II     TACAGTTACCANACACCACCTACCGGAGAGAAACAATTGTGGAGAATGGCC-CCTTATTT 936
type I      TACAGTTATCAAAGATGACCTACCAGAGAGAAACGATTGTGGAGAATGGCC-CCTTACTT 936
4p          ------TAGCAG-GA-GACCTGC---TGCG--------GTGGGGGCTGGGCTCCAGA--- 490
                  ** **   *  **** *    * *        **** *  *** * **  *

type II     TTGTTTATTTGCTGATNAGATTTCNTA-GTCCATTTC-TCATT--AGGTNCAGAGAT-CN 991
type I      TTGTNTATTTGCTGATAAGATTTCNTACGTCCATTTTATCATT--AGGTNCNAAGAT-CA 992
4p          ------ATATG--GGAAATACCTGCTACGTGAACGCT-TCCTTGGAGTGGCTGACATACA 541
                  ** **  *   * *  *  ** **  *     ** **  **   *  * ** *

type II     AAGTTGACCTACCCAAGGAGTTGA-GATATCCAGGNACNNAAA-CTCAGGGCACGGTAGA 1049
type I      AAGTTGACCTACCAAAG-AGTTGAAGATGTCCAGGA-CANNAA-CTCAGGGCTCAGTAGA 1049
4p          CACCGCCCCTTGCCAACTACATGC---TGTCCCGGG--AGCACTCTCAA---ACGTGTCA 593
             *     ***  * **  *  **    * *** **      *  ****     *     *

type II     ACCACAGAA-TCTTGGGTGAAATATTGCTCAAGAACAATAATGTNCTTATTCACC-GTGT 1107
type I      ACCNCNGAAATCTTGGGTGAAATATTGCTCAAGAACAAAAATGTNCTTATTCCCAANTGT 1109
4p          TCGTCACAA----GGGCTGCA----TGCTCT-GTACTA-----TGCAAGCTCACA----- 634
             *  *  **     ** ** *    *****  * ** *     * *    ** *

type II     TTCTGTGTGACATGTGTGGA-AACCAAAGTGCTTTGGAGCCTGACCCGAAGACTGAATTT 1166
type I      TTGTGTGTGACATGTGTGTTGAATTAAAGTGCAATG-AGCATGACATGCAGGCANGACAT 1168
4p          -------TCACACGGGCCCT---CCACAATCC--TG-GGCACGTCATCCAGCCCTCACAG 681
                  * *** * *         * * * *  **  **  * *    ** *   *

type II     CNTATTCGGCTC-CCTCCNNATCNGTTGTGTAGCCATTANAGAAAACCCAATTCCTAAGG 1225
type I      CN-ATTCGGCTCNCCTCAAAAGCAGTTNTGAA-CCTTAAAGGACAAC---AATCCTA-GG 1222
4p          -GCATT-GGCT----------GCTGGCTT----CCATAGAGGCAAGC------------ 712
               *** ****          * *   *    ** *    *  * *

type II     TCCCGCCTTAATGAANATTAGANCCATCCTTCCNCCTGTTGTNTAAANCCNCCGCATCTN 1285
type I      TCCCGC-TTAAAGAAA--TAAGACCATCCC-CCACCC--TGTGTTGAACCCCCGCNTCTG 1276
4p          ----------AGGAAG---ATGCCCAT-----------------GAA------------ 729
                     * ***    *   ****                  *

type II     GGGCTTGCTCATGNATTCCTGGGGGATCATTCCCCAAAAAAATGGGTGGGCTCCCTTCCT 1345
type I      GA--TTGCTCCTGT-TTCT--GGGGAACATCCTCCTGAAAA-TGGCGG---TCCTTTCC- 1326
4p          ----TTTCTCATGT--TCACTGTGGA----TGCCATGAAAAAGGCATG-----CCTTCC- 773
                ** *** **   **   * ***       *   ****  *   *     * ****

type II     CCCGG-- 1350
type I      CCCN--- 1330
4p          -CGGGCA 780
             *
```

Alignment b2:

```
type II     AACACACACAAACTGTTTAATATGAACACAATTGTTAAATAGCATTGTTATAACATGAAA 226
type I      CAAACACACAAACTGTTTAATATGAACACAATTGTTAAATAGCATTGTTATAT-ATGAAA 225
4p          CACACACGCACACGGTTTCCTAGGTAAAGATTTCTTCCCTGCCATTGCTTTAC-CTAAAA 4627
               **** ** ** ****  ** * * * * ** **   *  ***** * **   * ***

type II     TAAGGCAAATGTTTAGTTANTATCCTAACCCGGTTCCCATTCCTAACATATTTGCT-CAT 285
type I      TAAGGCAAATGTTTAGCTATTATCNTAGCCCG-TTCCCACACCTAACATATTTGCTTCAT 284
4p          TAAGGCAACTGTGAGGCCGCTGTCCCAACCCGGTTAC-ACTCCTATTATATGTGC----- 4682
            ******** ***   *    * **  * **** ** * *  ****  **** ***

Alignment b3:

type II     TTTAAAGGCGTTTGAAAGAAATTGAAACAGGAGAGTGGATGTATTCAGCTAAAATAAAAT 559
type I      TTTAAAAGAGTTTGAAAGAAATGGAAACAGAAGAGTGGATGTATTCAGCTAAAATAAAAT 558
4p          CTGTATTGATAAT-AAAGGAAAGCAAACACAGGAGTGTGTGTATTCAACTGAAATAAATT 2609
             *  *  *    * **** **   *****   *****  ******** ** ******* *

type II     CCGGAAGCCCTGAAATAAATNTCATTTTGCGGGTAAAAAAATGGCATTGGAGGAGATTCT 619
type I      CCGGAAGCCCTGAAATAAATNTCATTTTGGGTGTAAAAAAATGGCATTAGAGGAGATTCT 618
4p          CAGAAAGCCCTGAAATCAATCTCACTGGGTGTGTTTAAAAATGGCATTTGGGGAATTTCT 2669
            * * ************ *** *** *  * * **  ************ * ***  ****

type II     GGGTCAATCATG-AAGCTGTGAAAGTTGCATCTTGGAAGCA**GGATCC**CTGTAATGAAACG 678
type I      GGGTCAATCATCCAAGCTGTGAAAGCTGCATCTTGGAAGCA**GGATCC**CTGTAATGCAACG 678
4p          GGGTCATTTGTC-CAGCTGCGAAAGCTGCATCTCTGAAGCACAGTCCCTGTCCCGCAGTG 2728
            ****** *  *   ***** ***** *******  ******   *******   * *  *

type II     AGACTTGTTTATCAGAGGTGGTCTTTCAGANGAAAANATTTNGAAGAATGGCCCCTTCCT 738
type I      AGACTTGTATATCAGAGGTGGTCTTTCAGTGGAAAAGATTTTGAAGAATGGACCCTTCCT 738
4p          AGACTTATTTATCCGACGTGGTGTTTCCGTGGAAATGATTGTGGGAAATGGCCCCTTCCT 2788
            ****** * **** ** ***** **** *  ****  ***  *   ***** ********

type II     TTTGTGTATTTGACGATTAGATNTCATG-------------------------------- 766
type I      TTTGTGTATTTGACGATTAGACTTCATG-------------------------------- 766
4p          TTTCTCTATTTGCTGATTAGACTTCATGGTCCCTTTCTCGTCAGGTACAGTGATCAAAGT 2848
            *** * ******  *******  *****

Alignment b4:

type II     ATNTCATGCCA--AATTTC-------GGGTTTTAAACTCTATTTAANCNTTANAANAATT 809
type I      ACTTCATGCCA--AATCTC-------GGGTTTAAAACTCTATTTAAACATTAACAGAATT 809
4p          AAGCCCTGAAATCAATCTCACTGGGTGTGTTTAAAAATG-------GCATTTGGGGAATT 2666
            *   * **  *  *** **       * **** *** *         * **     ****

type II     ANC---TANAATGGCCAA-----AAAACAAGAAATTTTTTGATTAGGAATCGTCAAATAT 861
type I      AAT---TAAAATGGCAAA-----AAATCAAGAAATTTTTTGATTAGGAATCATCACATAT 861
4p          TCTGGGTCATTTGTCCAGCTGCGAAAGCTGCA---TCTCTGA-----AG----CACAG-T 3867
                  ** * *      *** *   *   * * ***     *     ** *  *
```

```
type II     TCATTTCTTG---TTAGATACAGTTACCANACACCACCTACCGGAGAGAAACAATTGTGG 918
type I      TCATTTCTTG---TTAGGTACAGTTATCAAAGATGACCTACCAGAGAGAAACGATTGTGG 918
4p          CCCTGTCCCGCAGTGAGACTTATTTATCCGACGTGGTGTTTCCGTG-GAAATGATTGTGG 3927
             * * **  *   * **    * *** *  *       *  * * * ****  *******

type II     AGAATGGCCCCTTATTTTTGTTTATTTGCTGATNAGATTTCNTA-GTCCATTTC-TCATT 976
type I      AGAATGGCCCCTTACTTTTGTNTATTTGCTGATAAGATTTCNTACGTCCATTTTATCATT 978
4p          GAAATGGCCCCTTCCTTTTCTCTATTTGCTGATTAGACTTCATG-GTCCCTTTC-TCGTC 3985
              **********  **** * ********** *** *** *  **** ***  ** *

type II     AGGTNCAGAGATCNAAGTTGACCTACCCAAGGAGTTGA--GATATCCAGGNACNNAAACT 1034
type I      AGGTNCNAAGATCAAAGTTGACCTACCAAAG-AGTTGAA-GATGTCCAGGA-CANNAACT 1035
4p          AGGTACAGTGATCAAAGTTGACCAACCCCAG-AG--GAAAGCTGCCCAGGG--CACAACT 4040
            **** *   **** ********* ***  ** **  **  * *  *****      ****

type II     CAGGGCACGGTAGAACCACAGAA-TCTTGGGTGAAATATTGCTCAAGAACAATAATGTNC 1093
type I      CAGGGCTCAGTAGAACCNCNGAAATCTTGGGTGAAATATTGCTCAAGAACAAAAATGTNC 1095
4p          CAGGGCTCCGTAGAACCACAGAA-TCTTGGGCGCAACCCTGCTCAAGCACCCAAATGTGC 3999
            ****** * ******** * *** ******* * **   ******** **   ***** *

type II     TTATTCACC-GTGTTTCTGTGTGACATGTGTGGA-AACCAAAGTGCTTTGGAGCCTGACC 1151
type I      TTATTCCCAANTGTTTGTGTGTGACATGTGTGTTGAATTAAAGTGCAATG-AGCATGACA 1154
4p          ATACGAACA-GGGTCTCCGTGTGACGTGTGTGAA-AACTACAGTGTGATG-AGCATGACT 4156
             **    *    ** *  ******* ******   **  * ****   ** *** ****

type II     CGAAGACTGAATTTCNTATTCGGCTC-CCTCCNNATCNGTT------------------- 1191
type I      TGCAGGCANGACATCN-ATTCGGCTCNCCTCAAAAGCAGTT------------------- 1194
4p          GGCAGACAGCTTATCG-ATTGGGCTCCCCTCAAAATCGGTTATGAGCATTCAAGCACACC 4215
             * ** *      **  *** ***** ****   * * ***

Alignment b5:

type II     CCGGAGAGAAACAATTGTGGAGAATGGCCCCTTATTTTTGTTTATTTGCTGATNAGATTT 959
type I      CCAGAGAGAAACGATTGTGGAGAATGGCCCCTTACTTTTGTNTATTTGCTGATAAGATTT 959
4p          TCCGTG-GAAATGATTGTGGGAAATGGCCCCTTCCTTTTCTCTATTTGCTGATTAGACTT 2812
             * * * ****  *******  ***********  **** * *********** *** **

type II     CNTA-GTCCATTTC-TCATTAGGTNCAGAGATCNAAGTTGACCTACCCAAGGAGTTGA-- 1015
type I      CNTACGTCCATTTTATCATTAGGTNCNAAGATCAAAGTTGACCTACCAAAG-AGTTGAA- 1017
4p          CATG-GTCCCTTTC-TCGTCAGGTACAGTGATCAAAGTTGACCAACCCCAG-AG--GAAA 2867
            * *  **** ***  ** * **** *   **** ********* ***  ** **  **

type II     GATATCCAGGNACNNAAACTCAGGGCACGGTAGAACCACAGAA-TCTTGGGTGAAATATT 1074
type I      GATGTCCAGGA-CANNAACTCAGGGCTCAGTAGAACCNCNGAAATCTTGGGTGAAATATT 1076
4p          GCTGCCCAGGG--CACAACTCAGGGCTCCGTAGAACCACAGAA-TCTTGGGCGCAACCCT 2924
            * *  *****      ********** * ******** * *** ******* * **   *

type II     GCTCAAGAACAATAATGTNCTTATTCACC-GTGTTTCTGTGTGACATGTGTGGA-AACCA 1132
type I      GCTCAAGAACAAAAATGTNCTTATTCCCAANTGTTTGTGTGTGACATGTGTGTTGAATTA 1136
4p          GCTCAAGCACCCAAATGTGCATACGAACA-GGGTCTCCGTGTGACGTGTGTGAA-AACTA 2982
            ******* **   ***** * **    *    ** *  ******* ******   **  *
```

```
type II     AAGTGCTTTGGAGCCTGACCCGAAGACTGAATTTCNTATTCGGCTC-CCTCCNNATCNGT 1191
type I      AAGTGCAATG-AGCATGACATGCAGGCANGACATCN-ATTCGGCTCNCCTCAAAAGCAGT 1194
4p          CAGTGTGATG-AGCATGACTGGCAGACAGCTTATCG-ATTGGGCTCCCCTCAAAATCGGT 3040
             ****   ** *** ****  * ** *      **  *** ***** ****   * * **

type II     TGTGTAGCCATTANAGAAAACCCAATTCCTAAGGTCCCGCCTTAATGAANATTAGANCCA 1251
type I      TNTGAA--CCTTAAAGGACAAC-AAT-CCTA-GGTCCCGCTTAAAGAAATA--AGA-CCA 1246
4p          TATGAG--CATTCAAGCACACC-GATGCCCA-GGTCCCGGCTGCAGGAATA--AGA-CCC 3093
            * **    * **  ** * * *  ** ** * *******     *     *  *** **

type II     TCCTTCCNCCTGTTGTNTAAANCCNCCGCATCTNGGGCTTGCTCATGNATTCCTGGGGGA 1311
type I      TCCC-CCACCC--TGTGTTGAACCCCCGCNTCTGGA--TTGCTCCTGTTT--CTGGGG-A 1298
4p          TCCA-GGGTCT--TGTGTGAAGCCTCGGCATCTGCA--TTGCTCATGCTT--CTGGGG-A 3145
            ***      *   *** *  * ** * *  ***     ****** **  *  ****** *

type II     TCATTCCCCAAAAAAATGGGTGGGCTCCCTTCCTCCCGG---------------------- 1350
type I      ACATCCTCCTGAAAA-TGGC--GG-TCCTTTC-CCCCN---------------------- 1331
4p          TCATTCTCCTGAAAA-TGGT--GGCTCCTTTC-TCCCTGTGGAGCATCTTTCTAAGCAGT 3201
             *** * **  **** ***   ** *** ***  *** *
```

c. Comparison of sequences C, c, and the 4p megasatellite.

```
type II     -------------------------AATTCTGTCCCTCTGAG----AATGCCGGA----- 26
type I      -------------------------AATTCTGTCCCTCTGAG----AATGCTGGA----- 26
4p          GTTGGGAAAATCCCTAGAGCCAGGATCTTCATTCCCGCTAAGCCAGACAGCCGGAAGACA 3529
                                      ***  **** ** **    *  ** ***

type II     --------------------------------CAGGTTTACCTTCATCACCATAA-AATT 53
type I      --------------------------------CAGGTTTACCTTCATCACCATAA-AATT 53
4p          CACCCAAATTCTGTCCCTCTTACTTCAGGGAACATGTCCACTTTCGGCAGCATTACAATT 3559
                                         ** **  ** ***  ** *** * ****

type II     TTGGAAC-AAATGTGGTAACTGCAGGTTCTCCCCACAATGAGTAACTGAAAATTGAGGCA 112
type I      TTGGAAC-AAATGTGGTAACTGCAGGTTCTCCCCACAATGAGTAACTGAAAATTGAGGCA 112
4p          TTGGCACCAAATGTGCTAACTGCAATTCCACCATACAATGCGTAACTGGAAATGGAGGCA 3619
            **** ** ******* ********  * * **  ****** ******* **** ******

type II     GTATTTCAGATCCTAAAAAACTGATGAAGTAATTCGCCACACATTTGGGTTGTTTTTGCC 172
type I      GTATTGCAGATCCTAAAAAACTGATGAAGTAATTCACCACACATTTGGGTTGTTTTTGCC 172
4p          ACATCTCCGATCCTGAACGATCGATGCGAGAATCCAGGATATGCACGGCTTATTTTGGCC 3679
              **  * ****** **  *  ****    *** *   * *     ** ** **** ***

type II     TTTTCCTACTATGAAGATTGCTAATAGGAAGAATAGTGTACAAGTATCTGTTTGATTCCC 232
type I      TTTTCCTACTATGAAGATTGCTAGTAGGAAGAATGGTGTACAAGTATCTGTTTGATTCCC 232
4p          TTTTCCCACTGAAACAAGGGCCAGTATTAAAAATGGCACGCTATCCTCTGTTTCACTCCC 3739
            ****** ***   *  *  ** * **  ** *** *    * *   ******* * ****

type II     TGCTTTTAGAATCCTTTGCTTGTTTGTGTTGTTTGTCTGTTCCTTCTTGAGACAGGATGT 292
type I      TGCTTTTAGAATCCTTTGCTTGTTTGTGT-GTTTGTCTGTTCCTTCTTGAGACAGGATTT 291
4p          TGCTTTTA-AAC---------GTCTCCGA--------TGTTTCTCCCTGAGACAGCGCCT 3781
            ******** **          ** *  *         **** ** * ********    *
```

```
type II     CACTCCCAGTCCAGCCCAGGATTTTCCCAGTT-TTGTTAATTTTTGTTGTTTCCCTTTTG 351
type I      CACTCCCAGTCCAGCCCAGGCTTTTCC-AGTTGTTGTTAATTTTTGTTGTTTCCCTTTTG 350
4p          CACTTCC-GTC-AGCCG-GGCTTTTCTACGGT----ATAATTTTCCTTGTTTGC-TTTTG 3833
            **** ** *** ****  ** *****   * *     *******  ****** * *****

type II     TTCAATTTTTAGAAATTTGTTATTTTATTTCCTATTGAATTTTAAGGCATTTTTAGATAT 411
type I      TTCAATTTTTAGAAATTTGTTATTTTACTTCCTATTGAATTTTAAGGCATTTTTAAATAT 410
4p          TCCAAAT--TAGAACTTT-TTATTTCATCTC-TAGGAAACGTTGATCCATTATCACATAC 3889
            * *** *  ***** *** ****** *  ** **   **  ** *  **** * * ***

type II     GTTATTAAAACATTATCCACACAGGCCGTGTTGTTTACATTGCAATTATTTCCACCA-TC 470
type I      TTT-TTAAAAANTTTTCCCCA-NNCCCGTNTT--TTAANTNGNAAT-ATCTCCCCCG-TC 464
4p          GTA-TGGAAATATTATC-ACACATGCTGTGAG--ATACGTTGTTTTTATTTTCATCAATT 3945
             *  *  ***  ** **  **    * **      **  * *   * ** * *  *  *

type II     CCCTTTAAGAAACCAAAGAAT----------TACCTTA---------------------- 498
type I      CC-TTNAAAAAACAAAGGTTT----------NACCTTA---------------------- 491
4p          CC--TTAATAAACAAAAGGTTATAGCTGGGATACCTTCTGAGTTCTCAAGTTTTTTGTTT 4003
            **  * ** **** ** *  *           *****
```

Appendix 2: BACs Isolated from the Research Genetics BAC library

Probes used for isolation: 87B23 T7 and SP6 ends. BAC clones isolated were in excess of 3,000. These BACs represent the strongest positive signals for clones present in the enriched chromosome 8 BAC library.

255E8
285P3
300F13
351I20
262F24
273J3
282B18
257E19
223B23
308D20
286D2
328H13
87G21
67B6
164L15
110A12

Note: 223B23, 67B6, 286D2 shown in a separate experiment to overlap with 87B23 but not contain the megasatellite