Positional Cloning Identifies A Candidate Gene For Mental Retardation And Epilepsy

Thesis submitted for the degree of Doctor of Philosophy

By

Kavita Bhalla M.Sc.

Department of Cytogenetics and Molecular Genetics,

Faculty of Medicine, Department of Paediatrics,

Women's and Children's Hospital,

University of Adelaide.

April, 2001

ERRATA FOR THESIS OF KAVITA BHALLA

Page 8	Section 1.3 line 3 replace "homologous chromosome" with "homologous chromosomes". Line 4 replace "the closer the two loci are" with "the closer that two loci are". Line 5 replace "will take place" with "will take place between them".
Page 9	Line 12, insert "and were highly informative" after "using PCR techniques".
Page 10	Line 13 replace sentence "the ability to positionally clone ....." "The ability to clone a gene based on a localization from a linkage study is dependant on the .....".
Page 12	Line 16 replace sentence "The STSs in the map ....." with "This STS-based map provided the technology that could be utilised to initiate large-scale sequencing of the human genome".
Page 17	Section 1.7 line 5 replace sentence starting "A transcript map ....." with "A transcript map is generated by identifying the contiguous and noncontiguous cDNAs or ESTs (which may represent different portions of a transcript) that show homology to the genomic sequence".
Page 18	Last paragraph replace "An alternative is to look ....." with "An alternative method for identifying genes is to look ....."
Page 19	Line 11 replace sentence "Kota and Sasaki (1998) used 2 o 4 base pairs restriction enzymes ....." with "To identify CpG islands, Kato and Sasaki (1998) used the restriction enzymes *Hpa*II and *Hha*I that cut frequently and specifically in CpG rich sequences instead of multiple rare-cutter enzymes".
Page 25	Line 8 replace ".....UniGene clusters representing human genes" with ".....UniGene clusters that are likely to represent genes, although in many cases there may be several UniGene clusters derived from the same gene".
Page 26/27	Replace sentence "In functional cloning, protein products, ..." with "Functional cloning involves the identification of the gene from knowledge of the protein using antibody screening of expression libraries or the use of protein sequence to identify homologous predicted protein sequence from nucleotide sequence".

Page 41	Line 3 replace "Unverricht-lundborg with "Unverricht-Lundborg".
Page 51	Section 2.2.5 replace "5g" with "5%".
Page 53	Replace "L. (Lauria) broth" with "L. (Luria) broth". Replace "Bacto-triptone" with "Bacto-tryptone".
Page 56	Line 1 replace "Bacto-triptone" with "Bacto-tryptone".
Page 63	Replace "2.3.2.1 Genomic DNA purification" with "2.3.2.1 Purification of cloned genomic DNA".
Page 68	Line 17 replace "3,000 rpm" with "3,000g"
Page 70	Heading 2.3.4 replace "gelelctrophoresis" with "gel electrophoresis".
Page 82	Line 13 replace "Digested bubble anchor – linkered DNA" with "Bubble anchor ligated, digested DNA".
Page 84	Line second from bottom. Insert "7.5ml" before "Express Hyb™ hybridisation buffer".
Page 148	Table 4.3. Captilisation of first base of sequence of reverse primer not necessary.
Fig. 5.4	Add to legend, "Time of electrophoresis run was 30 hours"
Fig. 5.8 and Fig. 5.9	In heading replace "poisitive" with "positive"
Page 204	Legend to fig. 6.2 line 8 insert "*Not*I and *Eag*I" after "using restriction enzymes".
Fig. 6.11	Line 8 of legend. Replace "lambda 4 *Hind*III" with "lambda DNA digested with *Hind*III."
Page 238	Line 11 replace "which is the closest" to "which was the closest".
Page 239	Line 15 replace "complaints" with "complain".
Page 242	Line 12 replace "β-ME" with "β-mercapto-ethanol".
Page 243	Line 6 insert "to the column" after "was added".

"It's all in the genes!"

N.G. Cooper

Los Alamos Science (20): 1992.

Table of contents

SUMMARY

Mental retardation and/or epilepsy account for the majority of childhood brain disorders. Genetic factors have an important role in these disorders. One approach to identify and characterize the genes involved in mental retardation and epilepsy is to identify the genes disrupted by the rare chromosomal rearrangements in patients having these phenotypes. Two unrelated patients were reported in the Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, Adelaide, with balanced *de novo* translocations t(1:16) and t(14;16) involved in mental retardation and epilepsy respectively. The mouse/human somatic cell hybrids isolated from the cell lines of these two patients containing the der(16)t(1;16) and the der(16)t(14;16) were designated as CY196 and CY182 respectively. The 16p13.3 breakpoints contained in CY196 and CY182 have been localized to the same physical interval on chromosome 16 at 16p13.3. It was hypothesized that the mental retardation and epilepsy in these two patients were due to the disruption of a gene at or near the translocation breakpoints at 16p13.3. This thesis mainly documents the detailed molecular characterization of the chromosome 16 breakpoints of these two *de novo* apparently balanced translocations. Positional cloning strategies were used to determine if any gene was disrupted at the breakpoints.

The current study began by generating a more detailed physical map of the breakpoints region in order to identify the disrupted transcript at 16p13.3. The preliminary physical mapping data identified a novel proline rich gene, C16*orf*5, highly expressed in the brain. Full-length cDNA characterization was performed and the genomic structure of the gene was established. Southern hybridization and FISH analyses localized the gene to chromosomal band 16p13.3, distal to the two

breakpoints. A YAC pulsed-field restriction map of the region was constructed to localize the C16*orf*5 gene in more detail. This map revealed that the C16*orf*5 gene was not directly disrupted by either of the translocation breakpoints at 16p13.3.

To further refine the two-translocation breakpoints at 16p13.3, a BAC/PAC contig was constructed across both the breakpoints. DNA of the PAC 978I1 spanning the 16p13.3 breakpoint of the t(1;16) was sequenced with 1.5X coverage to allow the possible identification of a transcript, which was disrupted at the breakpoint. Sequence for BAC 545E8 spanning the 16p13.3 breakpoint of the t(1;16) was available in GenBank. Sequence and expression analyses using the clones spanning the 16p13.3 breakpoints failed to identify any potential transcript.

FISH results of BAC 545E8 and a series of probes located proximal and distal to the 16p13.3 breakpoint of the t(14;16) suggested that the translocation in this patient was also associated with a paracentric inversion with a coincident breakpoint at 16p13.3, t(14;16)(p32;p13.3)inv16(p13.3p12.1). The unexpected three break complex chromosomal rearrangement complicated the further positional cloning approach.

The chromosome 1 breakpoint of the t(1;16) was shown to be within the pericentromeric heterochromatin. It was possible that the observed phenotype in the patient with t(1;16) might be the result of a position effect of the chromosome 1 heterochromatin. Expression and mutation analyses of the C16orf5 gene, the closest identified gene to the 16p13.3 breakpoint of this translocation were carried out. The data suggested that C16orf5 gene is not likely to be frequently involved in sporadic mental retardation.

Subsequently, the sequence of BAC 375G12, which was distal to the 16p13.3 breakpoint of t(1;16), identified a novel protein A2BP1 (ataxin–2 binding protein 1) (Shibata et al., 2000). A2BP1 cDNA sequence analysis together with the physical map data suggest that A2BP1 gene consists of 16 exons and spans over 1.6 Mb of genomic DNA. The chromosome 16p13.3 breakpoints of the t(1;16) and t(14;16) are located in introns of this gene. Since this brain and muscle specific transcript is disrupted by the 16p13.3 breakpoints of both *de novo* chromosomal rearrangements it is highly probable that A2BP1 is involved in the clinical phenotypes of the two patients. Identification of this disrupted gene provided a candidate for further mutation and expression analyses (currently underway in the Department of Cytogenetics and Molecular Genetics, WCH, Adelaide) in the other affected patients to assess the involvement of A2BP1 in sporadic autosomal mental retardation and/or epilepsy.

Declaration

This work contains no material that has been accepted for the award of any other degree or diploma in any University or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University library, being available for loan and photocopying.

Kavita Bhalla

Date: 27.4.2001

Publications

A list of paper and abstract published and manuscripts under preparation from the results presented in this thesis is given below. The copy of the published paper is provided in the appendix.

Papers:

Bhalla K, Eyre HJ, Whitmore SA, Sutherland GR and Callen DF (1999). C16*orf*5, a novel proline-rich gene at 16p13.3, is highly expressed in the brain. J Hum Genet **44**:383-387.

Callen DF, Eyre H, Schuffenhauer S and Bhalla K. A complex rearrangement involving simultaneous translocation and inversion demonstrates position effect compaction of chromatin. (manuscript in preparation).

Bhalla K, Eyre HJ, Kremmidiotis G, Gardner A, Crawford J, Gedeon AK, Mulley JC, Sutherland GR, and Callen DF. Two *de novo* balanced translocations associated with severe mental retardation and epilepsy disrupt a gene, A2BP1, at 16p13.3. (manuscript in preparation).

Abstract:

The American Society of Human Genetics Conference-2000 (Poster)

Bhalla K, Eyre H, Gardner A, Kremmidiotis G, Sutherland GR, and Callen DF (2000). Molecular analysis of severe mental retardation associated 16p-chromosomal breakpoint. Am J Hum Genet **67**(4) Supplement 2:165.

Abbreviations

ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
bp	base pairs
BLAST	basic local alignment search tool
C16*orf*5	chromosome 16 open reading frame 5
cDNA	complementary deoxyribonucleic acid
cM	centimorgan
DAPI	4',6-diamidino-2-phenylindole
dbEST	database of expressed sequence tags
DIG	Digoxigenin
dNTP	deoxynucleotide triphosphate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylene diamine tetra acetic acid
EST	expressed sequence tags
FISH	fluorescent *in situ* hybridization
FITC	fluorescein isothiocyanate
hnRNA	heteronuclear ribonucleic acid
kb	kilobase pairs
LANL	Los Alamos National Laboratory
Mb	megabase pairs
mg	milligram
ml	millilitre
mM	millimolar
mY	mega yeast artificial chromosome

µg	microgram
µl	microlitre
mRNA	messenger ribonucleic acid
NCBI	National Centre for Biotechnology Information
ng	nanograms
ORF	open reading frame
PAC	P1 artificial chromosome
PCR	polymerase chain reaction
PFGE	pulsed-field gel electrophoresis
RACE	rapid amplification of cDNA ends
RFLP	restriction fragment length polymorphism
RH	radiation hybrid
RNA	ribonucleic acid
RT-PCR	reverse transcription polymerase chain reaction
SSCP	single stranded conformation polymorphism
STS	sequence tagged sites
t	translocation
THC	tentative human consensus sequence
TRITC	tetra rhodamine isothiocyanate
UTR	untranslated region
VNTR	variable number of tandem repeats
v/v	volume per volume
WCH	Women's and Children's Hospital, Adelaide
w/v	weight per volume
YAC	yeast artificial chromosome
^{32}P	^{32}-Phosphorus

Acknowledgments

I would like to acknowledge the Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, University of Adelaide for providing the resources to perform the current research project. PhD fellowship support from AusAID is also acknowledged. Thanks are also due to the Department of Paediatrics, Faculty of Medicine, University of Adelaide for coordination of the PhD program.

I express my sincere gratitude to my supervisors, Associate Professor David Callen and Professor Grant Sutherland for their able guidance, warm support and sustained interest in my research work. Their timely and critical advise during the course of research and while writing thesis is acknowledged.

Special thanks are due to Rebecca Bilton, Joanna Crawford, Alison Gardner, Dr. Marina Kochetkova, Dr. Gabriel Kremmidiotis, Ingrid Lensink, Olivia McKenzie, Nathan O'Callaghan, Jason Powell, Genevieve Secker, Chatri Settasatian and Scott Whitmore for providing supportive working environment.

I greatly appreciate the timely help and support from Kathie Friend, Marie Hillman, Karen Lower, and Dr. Petter Stromme. I also acknowledge help from Elizabeth Baker and Helen Eyre in the FISH studies necessary for the project. Thanks are also due to Dr. Jozef Gecz, Dr. Oliva Handt, Lynne Hobson and Marie Mangelsdorf for their timely technical guidance. I also thank Dr. John Mulley and Dr. Agi Gedeon for providing DNA samples of the patients with non-specific mental retardation to carry out the mutation analysis of C16*orf*5 gene. My thanks are also to Dr. Norman Doggett, and Dr. Mei Wang for providing BAC clones, as

indicated in the text. I would like to thank Heather Cooke for her secretarial assistance. Personal thanks are due to all the Department members for providing a productive working environment during the course of this study.

I also take this opportunity to thank the University of Adelaide for providing me with Research Abroad Scholarship and the Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, Adelaide, for the financial support to be able to present the results from present study in the American Society of Human Genetics Conference, 2000, USA.

This is an opportunity for me to mention the inspirational support extended by my parents in my quest for higher education. The blessings from my in-laws were invaluable. I would like to acknowledge my brother in-law, Ravi and sister, Mausmi for their encouragement over the phone during some of the difficult times. Care from my brother Ateet has gone a long way in sustaining my interest to complete the present study.

Finally, and most importantly, I would like to acknowledge my husband, Shambhu, for the love and support during the course of this study.

This thesis is dedicated to my parents

Professor T.L. Bhalla and Mrs Kamlesh Bhalla

Literature Review

Chapter 1

1.1 Introduction

Over the years great progress has been made in unraveling the mysteries of human inheritance. It is now known that the human genome consists of 46 chromosomes, which between them house 3000 million base pairs of DNA. However, the likely number of genes present in the human genome is still not resolved. Early estimates suggested that there might be and 50,000 to 100,000 human genes (Schuler et al., 1996; Fields, 1994). Recently three reports were published, which suggests that the human gene number ranges from 30,000 (Ewing and Green, 2000; Crollius et al., 2000) to 120,000 (Liang et al., 2000). Alterations in these genes are responsible for nearly 4000 genetic diseases, which are known to afflict human beings (W1*). Before the age of 25 years, at least 53/1000 live borne individuals can be expected to have a disease with an important genetic component (Rimon et al., 1997).

The quest for an understanding of how genetic factors contribute to human disease is gathering speed. An international research effort is being carried out as a part of the Human Genome Project (HGP), which aims to characterize the human genome through complete mapping and sequencing of DNA, and to determine the estimated genes in the human genome (URL W1, Table 1.1). According to the current five year plan of HGP (1998-2003), the ultimate target to obtain the finished sequence of the human genome was expected to be achieved by 2003 (Collins et al., 1998). But the progress of the project today suggests that the final sequence of the human genome will likely be produced considerably ahead of that schedule. The HGP public and private consortium have already announced the

* Note: URLs, for www are listed in the Table 1.1 and referred as by the use of the W prefix.

working draft of human genome sequence on June 26, 2000, which can be accessed at URL W2, Table 1.1.

A private consortium of HGP, Celera Genomics (Rockville, MD) was started in May 1998, for the purpose of generating and commercializing genome information to accelerate the understanding of biological processes. Celera began to sequence the human genome on September 8, 1999 using the whole genome shotgun technique. The public and private projects use similar automation and sequencing technology, but different approaches for sequencing the human genome. The public project uses a 'hierarchical shotgun' approach in which individual large DNA fragments of known position are subjected to shotgun sequencing (that is, shredded into small fragments that are sequenced, and then reassembled on the basis of sequence overlaps). The Celera project uses a 'whole genome shotgun' approach, in which the entire genome is shredded into small fragments that are sequenced and put back together on the basis of sequence overlaps. Celera's human genome sequencing data has been submitted to the journal Science on December 6, 2000 and is expected to be accessible to the public by early 2001.

So far, two human chromosomes, 21 and 22, have been completely sequenced and published. Chromosome 22 was the first human chromosome to be sequenced. It constitutes 33.4 megabase of euchromatic DNA and contains at least 545 genes and 143 pseudogenes (Dunham et al., 1999). Chromosome 21 is the smallest human autosome and its sequence data (Hattori et al., 2000) revealed 127 known genes, 98 predicted genes and 59 pseudogenes. As of December 2000, 24.7% of

the human genome sequence is available as finished sequence (URL W3, Table 1.1).

Although the complete sequence of the human genome will soon be available, identification of the genes involved in diseases that have a complex pattern of inheritance, such as, mental retardation and epilepsy remains a challenge. Mutations in many genes related to mental retardation and epilepsy (section, 1.8) are known today. But due to the genetic heterogeneity associated with such complex diseases, etiologies of the majority of cases are still unknown. Further, there is no gene yet identified for a nonspecific form of autosomal mental retardation. Identification of yet unidentified genes related to mental retardation and epilepsy will provide more insights into the molecular mechanisms underlying these disorders.

One approach to the identification of such genes is to identify by positional cloning the genes disrupted at the breakpoints of chromosomal rearrangements associated with brain disorders. Chromosomal rearrangements associated with abnormal phenotypes have greatly assisted the positional cloning of many disease genes, examples are listed in Ballabio, (1993) and Collins, (1992, 1995). Two unrelated patients having *de novo* balanced translocations have been reported at the Department of Cytogenetics, Molecular Genetics, Women's and Children's Hospital, Adelaide. One of the patients had t(1;16) and severe mental retardation and the other had t(14;16) associated with grand mal seizures. Mouse/human somatic cell hybrids have been generated from the cell lines of these patients containing the der(16) of the t(1;16), designated as CY196 (Callen et al., 1990b), and the der(16) of the t(14;16), designated as CY182 (Callen et al., 1995). The

chromosomal 16 breakpoints in the two patients are located in the same physical interval of chromosome 16 at 16p13.3 (Callen et al., 1995). It was hypothesized that mental retardation and epilepsy in these two patients were due to a common gene defect at 16p13.3. Therefore, the present study aimed to identify the gene responsible for the phenotypes observed in the two patients using positional cloning strategies.

This literature review will discuss the background information available on human chromosome 16 as the current investigation focused the search for a gene responsible for the observed phenotypes in the two patients on chromosome 16. This will be followed by a discussion of the various genetic tools used for mapping and identification of disease genes. In the interest of providing updated information the discussion will include known genes responsible for mental retardation and epilepsy. Finally, general conclusions and the project aims will be presented.

1.2 Human chromosome 16

Human chromosome 16 harbors 98 Mb of DNA and represents approximately 3% of the human genome (Morton, 1991; Kozman et al., 1995). The expected number of genes on this chromosome is 1500-3000 (Mulley and Sutherland, 1993). The major initial objectives of the HGP included the construction of genetic and physical maps of each human chromosome to allow the cloning of disease genes linked to particular chromosomes. The early maps constructed were based on RFLP (restriction fragment length polymorphism) markers and were less efficient for positional cloning of disease genes than maps based on highly polymorphic PCR markers. Such maps are not discussed in this review.

Table 1.1 Uniform resource locators (URLs) for human genome information on the World Wide Web.

Web sites	URL address
W1	http://www.nhgri.nih.gov/HGP
W2	http://www.genome.cse.ucsc.edu
W3	http://www.ncbi.nlm.nih.gov/genome/seq/page
W4	http://www.ncbi.nlm.nih.gov/genome/guide/HsChr16.shtml
W5	http://www.ncbi.nlm.nih.gov/entrez/Omim
W6	http://www.ncbi.nlm.nih.gov/SNP/
W7	http://www.ornl.gov/hgmis/faq/snps.html#goals
W8	http://www.ncbi.nlm.nih.gov/genemap
W9	http://www-shgc.stanford.edu/RH/rhserver_form2.html
W10	http://www.sanger.ac.uk/Rhserver
W11	http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html
W12	http://www.ncbi.nlm.nih.gov/UniGene
W13	ftp://ftp.ncbi.nlm.nih.gov/repository/UniGene

There was a need to develop more informative maps which were helpful for positional cloning projects by placing more polymorphic markers based on dinucleotide repeats (which can be analyzed using PCR techniques) on the genetic linkage maps. Such a PCR based comprehensive genetic linkage map of chromosome 16 consisting of 79 AC repeats markers was reported by Shen et al., (1994). A year later, the integrated CEPH consortium genetic linkage of chromosome 16 was published by Kozman et al., (1995). The map contained 158 loci defined by 191 different probes. The mean genetic distance between markers was 2.8 cM. The markers were physically mapped on the human chromosome 16 somatic cell hybrid panel (Callen et al., 1992), which allowed the anchoring of the

genetic map to the cytogenetic map. These maps provided a scaffold for the construction of physical maps based on overlapping sets of DNA clones.

The PCR based genetic map and the CEPH consortium linkage map were merged with the physical maps by PCR screening of YAC and cosmid clones and assignment of markers to the hybrid breakpoint map (Doggett et al., 1995). This allowed the alignment of CEPH consortium genetic maps with the physical map. The physical map of chromosome16 consisted of a low-resolution megaYAC map and a higher resolution cosmid/contig/miniYAC map. The megaYAC map was constructed using 638 megaYACs from Centre d'Etude du Polymorphisme Humain (CEPH) that were localized to and ordered within the breakpoint intervals with 418 sequence-tagged sites (STSs). The megaYAC map provided almost complete coverage of the euchromatic arms of the chromosome 16. The high-resolution physical map consisted of 320 contiguous, overlapping sets of DNA clones (contigs) containing 2,000 fingerprinted cosmids and 248 miniYACs. This high-resolution physical map was integrated with the megaYAC map by clone hybridization and STSs screening. The high-resolution physical map covered over 50% of the euchromatic chromosome arms. This physical map also incorporated the integrated genetic and transcript map data generated by Callen et al., (1995) utilizing a high-resolution somatic cell hybrid panel. This included 141 genetic markers and 200 genes and transcripts. The integrated physical, genetic and cytogenetic map of chromosome 16 provided the basis for positional cloning of disease genes linked to this chromosome.

However, the integrated map mainly consisted of YAC and cosmid DNA clones (section, 1.4.2), which are no longer considered as substrates of choice for large-

scale DNA sequencing. Therefore, various groups then utilized the mapped and ordered STSs to construct BAC maps of chromosome 16 in order to allow large-scale DNA sequencing. As of December 2000, 21.7% of the chromosome 16 sequence is available as finished sequence (URL W4, Table 1.1). The chromosome 16 physical mapping data along with the ongoing sequencing efforts allows the rapid identification of candidate genes for diseases and traits linked to this chromosome. Chromosome 16 contains gene loci for several human diseases of both clinical and economic importance (Doggett et al., 1992). Currently there are nearly 269 entries that describe genes on chromosome 16 in the OMIM database (URL W5, Table 1.1).

1.2.1 A rodent/human somatic cell hybrid-based map of human chromosome 16

Somatic cell hybrid panels consisting of a mouse or hamster background and various portions of human chromosomes provide a basis for the rapid localization of genes or DNA fragments to individual chromosomes. A somatic cell hybrid panel was constructed for chromosome 16. Human cell lines containing chromosome 16 translocations and interstitial deletions were fused with the mouse cell line A9 and selection of the hybrids was possible utilizing the gene for adenine-phosphoribosyltransferase (APRT) located on the long arm of the human chromosome 16 (Callen, 1986). An expanded version of this panel is reported in Callen et al., (1990b), and Callen et al., (1992).

The hybrid panel was used for generating a high-resolution physical map of chromosome 16. Cloned DNA probes were localized to defined chromosomal intervals by Southern blot hybridization to DNA isolated from these somatic cell

hybrids. This panel was particularly useful in mapping smaller DNA probes (STSs) by PCR analysis of hybrid DNA (Callen et al., 1992). This high-resolution somatic hybrid panel for the entire chromosome 16 was further used for integration of physical, transcript and genetic maps of chromosome 16 (Callen et al., 1995; Doggett et al., 1995). The high-resolution cytogenetic based map of chromosome 16 divides the chromosome into 90 intervals of average size 1.1 Mb (Doggett et al., 1995). This cytogenetic based physical and transcript map of chromosome 16 provided the basis for cloning of the disrupted gene at the chromosomal 16 translocation breakpoints in the present study.

1.3 Genetic linkage mapping

The first step towards the gene identification is the assignment of a disease locus segregating in a family to a particular chromosome using genetic linkage analysis. Linkage analysis is based on crossing over between homologous chromosome during germ cell replication or meiosis. In principle, the closer the two loci are, the less likely it is that crossing over, observed as recombination, will take place. Two loci which show 1% recombination at meiosis are defined to be one centimorgan (cM) apart. A genetic distance of 1cM is approximately equal to a physical distance of 1megabase (Mb) averaged across the genome (Donis-Keller et al., 1987). A successful linkage analysis requires a family, or families, of adequate size in which a disease is segregating, to establish linkage with statistical significance. Another essential requirement for a linkage study is the availability of the set of polymorphic DNA markers with known location in the human genome, to assign a map position to the disease locus.

The first linkage map of the human genome was based on the pattern of inheritance of 403 polymorphic loci including RFLP's (restriction fragment length polymorphism) and VNTR's (variable number of short tandemly repeat DNA sequences) through a panel of 21 three generation families (Donis-Keller et al., 1987). However, this initial map was not highly informative and the average spacing between the markers was 10-20 cM. Subsequently, the HGP aimed to transform these existing limited genetic maps into a tool for finding disease genes by identifying and localizing highly informative polymorphic markers spaced at 2-5 centimorgans (cM) intervals along the genetic linkage map of each human chromosome (Hildebrand et al., 1992).

The most widely used DNA markers were microsatellites or short tandem repeats (STR) because they could be easily analyzed using PCR techniques. The repeated unit found in microsatellites is 2 to 6 base pair long, and they are abundant throughout the human genome (Weber and May, 1989). In 1993, a second-generation linkage map of the human genome based on such highly informative microsatellite loci was reported (Weissenbach, 1993). The use of simple sequence repeats as polymorphic markers greatly facilitated the construction of reference maps of the human chromosomes and mapping of Mendelian genetic diseases. The final updated version of the Genethon linkage map was published by Dib et al., 1996. This map consisted of 5,264 short tandem microsatellite repeat polymorphisms with an average heterozygosity of 70%. This map spanned a sex-averaged genetic distance of 3,699 cM and comprised 2,335 positions defined by 5,264 markers. The average interval size of this map was 1.6 cM with 59% of the map covered by intervals of 2 cM at most and 1% of the markers were greater than 10 cM apart. Such a map provided the basis for generating whole genome

physical maps based on overlapping set of yeast artificial chromosomes that were published by Chumakov et al., (1995) and Hudson et al., (1995).

Linkage analysis is an important and clinically applicable method of predicting inheritance of genetic diseases and traits. When a gene can not be detected directly, evaluation of a closely linked marker may be used to determine the likelihood that the gene has been inherited (Boughman et al., 1992). Linkage maps have proven to be extremely productive for mapping disease genes related to complex disorders such as, epilepsy (Delgado-Escueta et al., 1994), Alzheimer disease, Parkinson disease, Ataxia's, diseases of ion channels leading to periodic paralysis or hemilplegic migraine and tumor syndromes such as neurofibromatosis type1 and 2 (Pulst, 1999).

1.3.1 Single nucleotide polymorphism (SNP)

The ability to positionally clone a gene based on localization from linkage study is dependent on the requirement of a large family segregating the disease and a high resolution map of DNA-markers. Therefore, recent attention has focused on identification of SNPs (single nucleotide polymorphisms) as a source of genetic markers. SNPs are highly abundant and predicted to occur on average every 1 kb (Cooper et al., 1985). SNP is the most common type of human genetic variation and is a position in DNA at which two alternative bases occur at appreciable frequency (>1%) in the human population (Wang et al., 1998). In 1998, Lander and colleagues performed a survey to identify SNPs across the human genome using conventional gel-based DNA sequencing and high-density variation-detection DNA chips. A total of 3,241 candidate SNPs were identified in the 2.3 megabases of human genomic DNA examined, with 2,227 SNPs being used to

construct a genetic map. Subsequently, genotyping chips were synthesized that allow simultaneous genotyping of 500 SNPs, demonstrating the feasibility of large-scale screening of human SNPs (Wang et al., 1998). Currently, as of December 8, 2000, the SNP database (URL W6, Table 1.1) has submissions for 2,558,364 SNPs. Expectations are to have 150,000 SNPs mapped by mid-2001 (URL W7, Table 1.1).

SNP detection and scoring methods are many and various. Currently, most procedures involve target sequence PCR amplification, a burden that limits possibilities for scale-up and automation. This is equally true for the much-acclaimed miniature hybridization array (DNA-chip) concept. An effective method recently developed by the HGBASE (Human Genic Bi-Allelic SEquences) team is Dynamic Allele-Specific Hybridization (DASH) (Howell et al., 1999). Voltz et al., 2000 have already produced SNP maps for chromosome 21 and 22 and are working to extend their SNP data to other chromosomes as the finished sequence becomes available. A high-density SNP map is expected to help pinpoint genetic differences that predispose to disease and underlie variability in responses of individuals to treatment. Such maps will help to identify the multiple genes associated with complex diseases such as, cancer, diabetes, vascular disease, and some forms of mental illness.

1.4 Physical mapping

Physical maps of the human genome are essential tools for localization and positional cloning of disease genes. These maps consist of ordered, overlapping cloned DNA fragments covering each human chromosome. Physical maps have evolved over the past decade from the initial conception as a set of overlapping

clones (Olson et al., 1986) to the more recent idea of a well-spaced collection of unique landmarks (STSs) each defined by a polymerase chain reaction (PCR) assay (Olson et al., 1989). A first generation physical map of the human genome was published by Cohen et al., (1993). This map was prepared by screening a CEPH YAC library, which contained 33, 000 YAC clones, using 2,000 genetic markers randomly distributed over 90% of the human genome. This library had a ten-fold coverage of the haploid genome. Subsequently, a more detailed version of this map was published by Chumakov et al., (1995). This map covered 75% of the human genome and consisted of 225 YAC contigs with an average size of about 10 megabases. However, these maps were clone based, rather than STS-based. The YAC clones were not suitable for sequencing and STS coverage was too sparse to regenerate substantial physical coverage by use of other cloning vectors. Therefore, Hudson et al. (1995) constructed an STS based map of the human genome, which contained more than 15,000 loci with an average spacing of 199 kb. This map provided radiation hybrid coverage of 99 percent and physical coverage of 94 percent of the human genome. The STSs in the map provided a basis for initiating large-scale sequencing of the human genome.

With an advent of the HGP, an international consortium was formed to construct an integrated human gene map in which cDNA based STS markers were physically mapped and then integrated with the genetic map of polymorphic microsatellite markers (Dib et al., 1996). Schuler et al., (1996) published the initial report of this consortium that described a map of approximately 16,000 gene-based sequence tagged site markers that were mapped relative to a framework map which contained 1000 polymorphic genetic markers. This map unified the existing genetic and physical maps. This was further extended by

Deloukas et al., (1998) to include 30,000 human gene-based markers. Within two years, this map represented nearly a 100% increase in gene density and map accuracy and was now estimated to contain half of all human protein-coding genes. This integrated map can be accessed at URL W8, Table 1.1.

These physical maps are now being replaced by the DNA sequence of the human genome. An update of the information on the ongoing mapping and sequencing of different human chromosomes, including human chromosomes 21 and 22 whose genomic sequence has been completed, can be accessed at URL W3, Table 1.1

1.4.1 Radiation hybrid (RH) mapping

The STS-based physical maps of chromosome 16 (section 1.2) reported by Doggett et al., (1995) as well as the entire human genome maps of Cohen et al., (1993), Chumakov et al., (1995) and Hudson et al., (1995) (section 1.4) used YAC DNA clones for integrating maps. However, certain GC-rich regions of the human genome appear to be poorly represented in YAC libraries (Chumakov et al., 1995). Therefore, a radiation hybrid (RH) mapping approach was utilized to order STSs and integrate them with the genetic maps of polymorphic markers. RH panels consist of hamster cell lines that contain many large fragments of human DNA produced by radiation breakage (Goss and Harris, 1975). RH mapping, unlike linkage mapping which is limited to the use of polymorphic markers allows mapping of any unique non-polymorphic markers such as STSs and ESTs. The markers can be mapped on the RH panel by PCR or Southern analyses of DNA isolated from the respective radiation hybrids.

At present, three different hamster-human whole genome RH panels have been constructed and are available (all distributed by Research Genetics, Huntsville, AL) to the scientific community. GB4 RH panel consisted of 93-hamster cell lines, each consisting of the human genome in random fragments of approximately 10 Mb (Gyapay et al., 1996). This panel was used by Genethon and Cambridge University to construct an RH map of the human genome using 850 STSs and markers from genetic linkage map (Gyapay et al., 1996). The GB4 RH panel was constructed using a low dose of irradiation (3000 rads of X-rays), which allows the construction of continuous maps when these hybrids are scored with as few as 400 STSs spanning the human genome. However, this immediate advantage limits the ultimate resolution of the maps that can be generated with these reagents.

The G3 RH panel was generated using a relatively high dose of irradiation (10,000 rads) for the construction of RH panels of high-resolution. The G3 RH panel consisted of 83 hamster cell lines, each retaining approximately 16% of the human genome in random fragments of approximately 2.4 Mb (Stewart et al., 1997). Both the GB4 and G3 RH panels were used by Schuler et al., (1996) to map gene-based sequence tagged site markers relative to RH and YAC panels. The third panel, the TNG RH panel was generated at the Stanford Human Genome Center with 50,000 rads of irradiation and can be used to order markers at 50 kb resolution (Lunetta et al., 1996). These three different RH panels allow the ordering of markers at different levels of resolution. The GB4 RH mapping data were merged and aligned with the G3 RH map and the genetic map (Dib et al., 1996) utilizing 1102 microsatellite markers that were common to all three maps. The existing genetic and RH maps span 100% of the human genome

(Schuler et al., 1996). The integrated map is available at URL W8 Table 1.1. In addition, two web servers, one for each RH panel (G3 RH panel =URL W9, GB4 RH panel =URL W10, Table 1.1), allow mapping of any new marker relative to this map.

In the present study, the information on STSs and ESTs present in the vicinity of the two breakpoints was obtained from existing integrated CEPH genetic and physical map and RH map for chromosome 16.

1.4.2 Cloning Systems

To facilitate the ongoing efforts to map and sequence the human genome and to isolate the disease genes, it is important to clone chromosomal fragments as a single contiguous piece of DNA. Cosmid and the yeast artificial chromosome (YAC) cloning systems have the potential to accomplish these goals but both have demonstrated deficiencies. The yeast artificial chromosome (YAC) cloning system has permitted the cloning of large fragments of mammalian DNA (Burke et al., 1987). The YAC cloning system can accommodate genomic inserts that are up to or even greater than 1 Mb in size. However, the major disadvantage of the YAC cloning system is that a large percentage of clones, up to 50%, contain DNA segments from non-contiguous parts of the genome, that is, are chimeric clones. YAC cloning also appears to be prone to producing deletions and/or rearrangements of the cloned insert, which restricts their reliability for mapping and sequencing purposes (Shepherd et al., 1994; Boycott et al., 1998). Cosmids have an upper insert size limit of only 45 kb (de Jong et al., 1989), making the process of chromosome walking a laborious and inefficient effort (Boycott et al., 1998). Genomic sequencing based on cosmids is also inefficient as shot gun

approaches will result in maximum percentage of the sequence generated being vector sequence, which is discarded.

In recent years, other bacterial vector hosts systems, which have clonal stability and larger insert sizes have been developed. These are considered as the vectors of choice for the process of chromosome walking and DNA sequencing. These cloning systems include PACs (P1-derived artificial chromosomes) and BACs (bacterial artificial chromosomes). PAC clones have average insert sizes of 100 kb. The PAC cloning system is based on bacteriophage P1 (Sternberg, 1990) and allows rapid isolation of recombinant DNA. Because of the clonal stability of PAC clones, PAC libraries have a faithful representation of the genomic DNA. Similar to cosmids, the inserts of PAC clones are also flanked by rare-cutting restriction enzymes as well as by T7 and SP6 promoters (Shepherd et al., 1994), which allow the direct sequencing of PAC ends. However, in this system, human DNA inserts greater than 120 kb have not been cloned and characterized, which limits the use of this system.

The size limitation of PACs were circumvented by development of the BAC cloning system, which allows the cloning of large genomic inserts greater than 300 kb. This system is based on the well-studied *Esherichia coli* F factor. The F plasmid is maintained in low copy number (one or two copies per cell), thus reducing the potential for recombination between DNA fragments carried by the plasmid (Shizuya et al., 1992). However, the low copy plasmid number makes it difficult to obtain large amounts of DNA from bacterial cultures as compared with PACs. Inserts in BAC clones are also flanked by T7 and SP6 promoters, which can be used to design primers for generating BAC end sequences directly from

BAC DNA. These sequences can then be used to design end probes for another cycle of library screening in order to extend the existing BAC contigs. Due to the large insert size and clonal stability the BAC system is now considered as the vector of choice for generating large-scale DNA sequence.

1.5 Transcript mapping

Of the $3X10^9$ bases of the human genome only a small percentage codes for proteins, and coding regions are generally fragmented into exons. Therefore, the identification of coding regions (genes) in the human genome is a major challenge. However, this can be facilitated by the construction of transcript maps. A transcript map is generated by identifying contiguous and noncontiguous cDNAs or multiple ESTs (which represent different portions of a trancript) in the implicated region. The finished sequence of the human genome will soon be available but the biological significance of most of the sequence is unknown. Therefore, a transcript map of the human genome will be a critical resource for defining the regions that code for proteins. Over the years, various methods of transcript identification have been developed and these will be discussed.

The first whole genome transcript maps were constructed in the mid to late 1970's by analysis of DNA hybridization on viral or organelle genomes by means of S1 nuclease mapping or electron microscopy (Battey and Clayton, 1978; Sharp et al., 1980). Because of the small size of these genomes, these techniques allowed characterization of the temporal aspects of gene expression as well as gene locations. However, these methods were not suitable for constructing human gene maps.

In the 1980s one method commonly used for the identification of coding regions in the human genome involved the screening of short genomic DNA segments for sequences that were evolutionary conserved. For this, genomic DNA was isolated and used as a probe on Southern filters containing DNA isolated from a number of different species (Zoo blots) to test for evolutionary conservation of a particular DNA segment (Monaco et al., 1986; Rommens et al., 1989).

A different approach was used by Fearon et al., (1990) where they analyzed the genomic DNA for the presence of open reading frames. To determine, if the genomic DNA of interest is a part of the gene they probed a Northern blot-containing RNA obtained from different human tissues. If a transcript was detected the probe was subsequently used to screen the cDNA library (random collection of cloned cDNAs) generated from the relevant tissue(s) to identify a corresponding cDNA clone. The clone was then sequenced and analyzed for the presence of open reading frames. These methods were used in the identification of transcripts involved in Duchenne muscular dystrophy (Monaco et al., 1986), Cystic fibrosis (Rommens et al., 1989) and DCC gene responsible for colorectal cancers (Fearon et al., 1990).

An alternative is to look for CpG islands, which can serve as a useful gene marker (Bickmore and Bird, 1992). CpG islands are genomic regions larger than 200 basepairs with a high G + C content (greater than 50%) (Gardiner-Garden and Frommer, 1987). Although cytosine residues of most CpGs in the genome are partially or completely methylated, those in CpG islands are normally free from methylation. CpG islands are usually associated with the promoter and the 5' portion of a gene (Bird, 1987). It has been estimated that there are 45,000

unmethylated CpG islands in the human genome, and 56% of all human genes are associated with CpG islands (Antequera and Bird, 1993). The most common approach used to identify CpG islands in a defined genomic region is to map the cleavage sites for rare-cutter restriction enzymes. Rare-cutter restriction endonucleases recognize 6 to 8 bp sequences containing at least one CpG dinucleotide (Bickmore and Bird, 1992). Therefore, if several rare-cutter sites are co-mapped in a defined region, the region is most likely to be a CpG island. However, Bickmore and Bird, (1992) found that the frequency of rare-cutter sites in CpG islands averages 0.27-2.14 sites per island, based on an average island size of 1.4 kb. Therefore, one could overlook a proportion of islands by this strategy.

Kato and Sasaki (1998) used 2 to 4 base pair restriction enzymes such as *Hpa* II (CCGG) and *Hha* I (GCGC) instead of multiple rare-cutter enzymes to identify CpG islands. Multiple sites are present for both *Hpa* II and *Hha* I in CpG islands (average 20.35 and 21.86 sites per island, respectively) and their density is approximately one in every 70 base pairs per island and one per 2300 base pairs in the remainder of the genome (Bickmore and Bird, 1992). This large difference in frequency enabled them (Kato and Sasaki, 1998) to detect CpG islands as clusters of sub-bands after partial digestion of large DNA fragments with *Hpa* II and *Hha* I restriction enzymes, and subsequent Southern hybridization with end probes generated from cosmid and P1 clones. However, there are islands not associated with genes. For example, many transposable elements and pseudogenes are known to have CpG islands, some of which may be unmethylated, and as some tissue-specific genes have one or more islands associated with cryptic internal promoters that do not produce translatable transcripts (Larsen et al., 1992).

Therefore, this method does not provide a direct means of detecting coding sequence from the genomic DNA but is useful in pinpointing the possible genes.

Use of a hn (heteronuclear) cDNA library (Liu et al., 1989, 1992) is yet an another method for the isolation of coding regions from genomic DNA. The hn-cDNA libraries are made from unprocessed messenger RNA by using a primer that binds to 5' intron consensus splice site sequences to initiate cDNA synthesis (Liu et al., 1989; Whitmore et al., 1994). Chromosome specific cDNA libraries can be generated using hn-cDNA derived from somatic cell hybrids containing the chromosome of interest as the only human DNA content. Liu et al., (1989, 1992) utilized this approach to generate a hn-cDNA library from a somatic cell hybrid containing the long arm of chromosome 19.

Alternatively, Corbo et al., (1990) used conserved regions of human Alu repeat sequences to design oligonucleotides for generating cDNA from hnRNA. The resulting library was screened with human specific repeats to identify human clones. A similar approach was used by Whitmore et al., (1994) to generate a chromosome 16 hn-cDNA library from the somatic cell hybrid CY18, which contained chromosome 16 as the only human chromosome. Advantages of this approach were that cDNA synthesis was initiated through an exon, and therefore each clone isolated was expected to contain coding sequence. Isolation of cDNA from the somatic cell hybrids also provides the knowledge of their chromosome location. In addition, this approach also helps in the isolation of cDNAs that may be in low abundance in those libraries derived from particular tissues. However, this approach is tedious and results in isolation of large numbers of rodent clones due to non-specific priming events.

A number of researchers also used genomic DNA such as YAC inserts to directly screen cDNA libraries involving Southern hybridization techniques and blocking of repeat sequences in the large inserts prior to hybridization (Elvin et al., 1990; Geraghty et al., 1993). However, one of the major problems was that due to the large size YAC insert the smaller cDNA clones present in the library remained undetected. To overcome this problem cDNA clones within the libraries were first enriched using a cDNA selection method (Parimoo et al., 1991) before being screened for hybridization with probes containing large YAC inserts.

The cDNA selection approach involved the immobilization of target cosmid or YAC DNA to nylon filter discs. The prepared discs were then blocked for repetitive, ribosomal or GC-rich sequences and hybridized with cDNA fragments prepared by PCR amplification from a λgt10 cDNA library using flanking vector primers. After an appropriate wash, the hybrid-selected material was PCR-amplified, analyzed on Southern blot and cloned in λgt10. Using this approach (Parimoo et al., 1991) were able to identify several non-major histo compatibility complex class I clones from a yeast artificial chromosome that includes the HLA-A locus. This strategy had several advantages over other approaches discussed above. This method was able to detect 5' and 3' exons and genes that lack CpG islands and those genes that are not expressed in the hybrid cell lines. This method was particularly convenient for YACs where a single gel and blot could be used to identify cDNAs corresponding to several YACs in a short period of time. However, the limitation of this approach is that the recovery of genes is dependent on the presence of the transcript in the cDNA library. Rare transcripts or those that are expressed in a tissue specific manner will be difficult to identify.

The cDNA selection approach was further modified by Korn et al., (1992). The method involved the hybridization in a solution (rather than filter-bound templates), of amplified cDNA inserts to biotinylated cosmid DNA. The selected cDNA clones were then "captured" using streptavidin coated magnetic beads, which can be isolated by passing over a magnet, while non-selected cDNA clones were washed from the hybridization. The eluted cDNA were reamplified by PCR, cloned and analyzed further. Candidate genes for Huntington's disease (Rommens et al., 1993) were identified using this method. However, this method results in preferential amplification of abundant cDNA clones in a given library resulting in under representation of the rarer transcripts. Further, medium to low copy repeat sequences are difficult to suppress and this usually results in the screening of a high frequency of false positive clones.

An exon amplification strategy based on RNA splicing was developed to tackle the problems observed with the existing methods to identify human genes. This method is based on the direct functional detection of exon sequences in genomic DNA by recognition of flanking functional 5' and 3' splice sites. Fragments of cloned genomic DNA are inserted into an intron, which is flanked by 5' and 3' splice sites of the human immunodefeciency virus 1 *tat* gene contained within the plasmid pSPL1. These constructs are then transfected into COS-7 mammalian cells where the genomic introns are spliced out resulting in mature RNA. The resulting mature RNA contains the trapped exons, which can be amplified by RT-PCR and cloned for further sequence analysis of the trapped exons (Buckler et al., 1991). 50% of the exons trapped using the pSPL1 vector contained sequences derived from the intron of the trapping vector. This was due to the presence of cryptic splice sites in the cloning site of pSPL1 trapping vector. Therefore, to

increase the exon trapping efficiency (Burn et al., 1995) designed another trapping vector called pSPL3B. They found that this vector eliminated the pSPL3B only spliced products. The pSPL3B-based system was 98% efficient in the ability to trap exons in genomic DNA. One disadvantage of this approach is that the trapped exons are usually small in size and are therefore not suitable as probes for Northern blot analysis. Therefore, confirmation of the trapped exons as a part of a gene requires that adjacent exons be linked by RT-PCR. This process of gene identification is tedious and time consuming.

The identification of human genes was revolutionized by the advent of large-scale sequencing. This gradually transformed tedious laboratory based methodology into largely in silico based approaches. Brenner (1990) proposed the idea of large-scale sequencing of cDNA as a part of the HGP to identify the complete set of human genes. Adams et al., (1991) focused on generating short cDNA fragments, which they termed as "expressed sequence tags" (ESTs). The EST strategy was developed to allow rapid identification of expressed genes by sequence analysis. With the advent of high-throughout cDNA sequencing, a special division of the GenBank sequence database was created in 1992 (Boguski et al., 1993), designated as dbEST, a database of EST sequences.

The majority of ESTs are derived from cDNAs that are directionally cloned from their 3' ends using an oligo (dt) primer that anneals to the polyadenylate (poly A) tail found at the end of most human mRNA (Wilcox et al., 1991). The 3' ends of the cDNA sequences are used to generate EST sequences because 3'UTRs (untranslated region) show less sequence conservation than the coding regions of the gene. This makes it easier to discriminate among gene family members that

are similar in their coding regions. In addition the 3' UTRs are generally intronless, which enables the design of oligo primers from the EST sequences for PCR amplification of expressed sequences from the genomic DNA.

In an effort to identify new genes and analyze their expression patterns, Adams et al., (1995) generated 174,472 EST sequences from cDNA libraries constructed from 37 distinct organs and tissues. This data added up to 52 million nucleotides of human DNA sequences. Since then, there has been a continued increase in the collection of EST sequences in dbEST. As of December 25, 2000, dbEST (dbEST release 121500) contains 2,816,160 human EST sequences (URL W11, Table 1.1).

By 1995, the number of the ESTs in the GeneBank was increased by thousands and therefore it was necessary to create a human gene catalogue. A gene is represented by multiple ESTs, which may correspond to different portions of a transcript. So to reduce the extent of redundancy and to build longer contiguous blocks of sequence, Adams et al., (1995) assembled 118,406 ESTs, present at that time in dbEST, into 29,599 distinct assemblies (contigs). These assemblies are known as THCs (tentative human consensus sequences), and were notated with information on source library and abundance of ESTs and in some cases represent full-length cDNA transcripts. In 1996, Schuler et al., developed an information resource called UniGene (URL W12, Table 1.1), where EST data are represented in such a way that all representations of a single gene are collected in a single UniGene cluster. Schuler et al., (1996) arranged their data of 163,215 3' EST sequences together with 8516 3' ends of genes known at that stage into 49,625 clusters. Of these, 4,563 (9%) corresponded to known genes, and the remainder

was represented only by ESTs. The HGP aimed to localize these UniGene clusters on the RH and YAC physical maps of the human genome. This was done to accelerate the positional cloning of human genes. The last version of the human gene map published in 1998 by Bentley and colleagues resulted in the placement of 18,703 UniGene clusters on the physical map of 30,000 human genes. The remaining unmapped and newly identified UniGene clusters and cDNAs were progressively mapped after this time. As of December 25, 2000, there are 86,513 human UniGene clusters representing human genes (URL W13, Table 1.1).

With the available EST and cDNA sequence data in dbEST, the present method of choice to identify genes is to sequence the genomic insert of interest and then analyze the sequence data for the presence ESTs and genes using publicly available sequence comparison tools. BLAST (Altschul et al., 1990, 1994) and FASTA (Wilbur and Lipman, 1983) are the commonly used NCBI (National Center for Biotechnology Information) search tools for sequence homology searches.

The most recent developing approach is cDNA microarray technology or DNA chip technology. The technology is based on an approach where known human genes are roboticallly printed onto a glass slide and these arrays are subsequently hybridized to fluorescent labelled pools of cDNA. The cDNAs are generated after isolating mRNA from cells or tissues that one wishes to study. Microarray technology allows expression monitoring of hundreds and thousands of genes simultaneously, and provides a format to identify genes as well as changes in their activity (Heller et al., 1997). The ultimate goal of this technology is to develop the arrays that contain every gene on the human genome against which mRNA

expression can be assessed. This technology has been recently developed and therefore could not be used in this thesis. Further, an array with a complete set of human genes is not yet available and this is essential for a positional cloning strategy based on this approach.

In the current study, the following EST based strategy was chosen as an approach to positionally clone the disease gene, as this is a rapid and efficient method of gene identification. The breakpoints at 16p13.3 will first be localized to a BAC or PAC and subsequently the genomic clone of interest will be sequenced. The sequence data generated will be analyzed for the presence of ESTs and candidate genes in dbEST using the BLAST algorithm at NCBI. The predicted ESTs and cDNAs will be probed to a commercially obtained Northern blot in order to identify the size and tissue expression of the corresponding transcript.

1.6 Cloning of human disease genes

The detailed genetic, physical and transcript maps of the human genome have greatly facilitated the cloning of disease genes that contribute to the pathophysiological mechanisms involved in a clinical phenotype. In general, the disease genes can be cloned using functional cloning, positional cloning and positional candidate approaches, which are discussed below.

1.6.1 Functional cloning

In the past, a large number of disease genes were identified on the basis of the pre-existing knowledge about the basic biochemical defect (examples, phenylketonuria and sickle cell anemia). This approach was referred to as "functional cloning" (Collins, 1992). In functional cloning, protein products,

partial aminoacid sequences and antibodies against the protein products are mainly used as probes to screen the cDNA libraries in order to isolate the disease genes. While this approach of functional cloning was remarkably successful, it was limited to disorders where functional information about the basic defect exists. Unfortunately, for large number of monogenic Mendelian disorders this biological information is not known. Therefore, this route to gene identification is not commonly used.

1.6.2 Positional cloning

Disease genes can be isolated solely on the basis of map position without any knowledge of gene function using positional cloning (Collins 1992, 1995). This process of gene identification by map position was originally referred to as reverse genetics (Ruddle, 1984; Orkin et al., 1986). The positional cloning of the disease genes is initiated from the knowledge of genetic/physical location of a disease locus in the genome. This is followed by the successive narrowing of the interval using any of the cloning systems discussed in section 1.4.2. Subsequently the transcript map of the implicated region results in identification of the disease gene(s), whose function can then be studied. The experimental steps involved in the positional cloning of a disease gene(s) are illustrated in Figure 1.1. The first success in positional cloning was reported in 1986 with the cloning of the X-linked gene for chronic granulomatous disease (Royer-Pokora et al., 1986). Since then, many genes for inherited disorders are identified using positional cloning strategies, examples are listed in Ballabio, (1993) and Collins (1992, 1995).

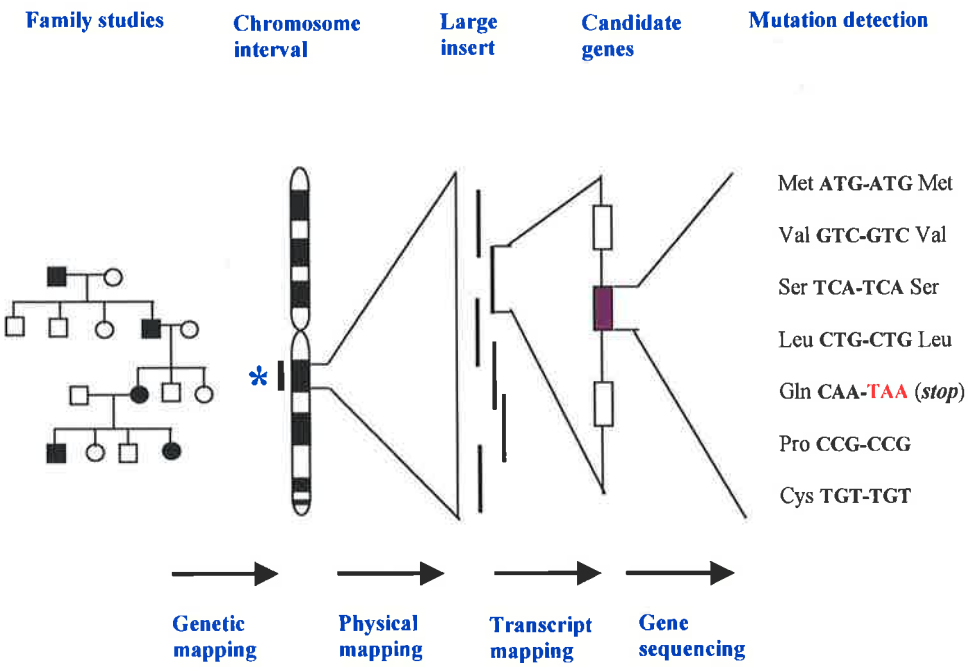


Figure 1.1 Steps involved in the positional cloning of the disease gene(s)

(Adapted from, Schuler et al., 1996)

(The positional cloning strategy will be used by the candidate to identify the disrupted gene(s) at the 16p13.3-translocation breakpoints in the two patients having mental retardation and epilepsy).

1.6.3 Positional candidate approach

The first step towards the gene identification is the localization of a disorder to a particular chromosome using genetic linkage analysis. However, if large families are unavailable for a precise genetic localization, the genetic mapping is unlikely to allow the localization to less than a few megabases. Identification and screening of all the genes from large linkage intervals (1 gene/30 kilobases) (Ballabio, 1993) will be time consuming and unproductive. In such a case, a modification of the approach, the positional candidate approach (Collins, 1992; Ballabio, 1993) of gene identification may be appropriate. This approach relies on the combination of information on the function and position of a particular gene. Due to the efforts of HGP, there is a steady increase in the transcript density for each human chromosome. Therefore, when a particular disorder is assigned a map position, it is now possible to access a list of all the genes assigned to that region. These genes are then considered as candidates for that disorder. The features of these genes are compared to the disease phenotype, in order to find strong candidate gene(s), which may have a role in that particular disorder. This approach has been successful in identification of many genes (reviewed by Ballabio, 1993 and Collins 1992, 1995). The success of this approach depends upon the availability of the dense transcript map and functional information about the disease. In future, with the availability of the finished human genome sequence and the complete transcript map of all the estimated human disease genes, the positional candidate approach will probably become a method of choice.

1.7 Chromosomal rearrangements: A unique resource for disease gene identification

Chromosomal translocation results from the reciprocal exchange of chromosomal segments between non-homologous autosomes. Balanced chromosomal rearrangements of the autosomes have been observed in 0.15% of live born human infants (Tharapel et al., 1977). While the majority of chromosomal translocations are inherited, some originate *de novo*. The incidence of balanced *de novo* translocation is about 0.02% as calculated from the newborn surveys (Tharapel et al., 1977). In general, carriers of balanced chromosomal translocations are associated with normal phenotype, assuming that the rearrangement involves no gain or loss of the chromosome material. However, the risk of serious congenital anomaly is estimated to be 6.1% (n=163) for *de novo* reciprocal translocations (Warburton, 1991). This study involves the detailed molecular characterization of the de novo balanced translocations t(1;16) and t(14;16) in two patients with mental retardation and epilepsy respectively. Such translocation cases provide a unique resource for a positional cloning strategy leading to identification of genes at or near the chromosomal breakpoint, the functions of which may be disrupted by the rearrangement and which may be responsible for the disease phenotype observed (Bedell et al., 1996; Baere et al., 2000).

Positional cloning of a disease gene begins with the mapping of a disease locus to a particular chromosome using linkage analysis. For a precise genetic localization of a disease locus, large pedigrees and highly informative markers are required. In practice, the availability of large families is not always possible, so the fine mapping of disease locus is often limited to large genetic intervals. This makes

isolation of a disease gene from the implicated region tedious using a positional cloning strategy or the positional candidate gene approaches. A possible alternative is the positional cloning of chromosomal rearrangements associated with clinical phenotypes (Villard et al., 1999). However, the occurrence of such cases is very rare, and therefore when identified, it is important to explore them for the identification of disease genes. The availability of patients with chromosomal rearrangements can greatly accelerate the process of disease gene identification (Collins, 1992; 1995).

The existence of chromosomal rearrangements associated with abnormal phenotypes has greatly assisted the positional cloning of disease causing genes. For example, a translocation with one breakpoint on the long arm of chromosome 17 in an individual with neurofibromatosis-1 provided an entry to cloning of the NF-1 gene (Ledbetter et al., 1989). Another important example was the identification of the gene for Duchenne muscular dystrophy through the study of a single patient with concomitant muscular dystrophy, McLeod syndrome and chronic granulomatous disease. The patient had an X chromosome deletion, which had removed a set of contiguous genes (Koenig et al., 1987). Cloning of most of the common chromosomal breakpoints involved in lymphoma and leukemia have also led to the identification of disease associated genes (Collins, 1995). This thesis involves the molecular analysis of two translocation breakpoints at chromosome 16p13.3 associated with mental retardation and epilepsy to identify the disease related gene.

1.8 Disease genes involved in mental retardation and epilepsy

Mental retardation and/or epilepsy account for a majority of childhood neurodevelopmental disorders. Therefore, a major priority of neuroscience research is identification of genes related to brain disorders. Of all the estimated genes expressed in man, it appears that over half have some role in brain structure and function (Adams et al., 1995). Increased number of genes with known and unknown functions are being identified and localised as a part of the HGP. Identification of genes related to mental retardation and epilepsy will not only allow researchers a better understanding of the molecular basis of neuronal mechanisms but will also help clinicians in presymptomatic diagnosis and designing rational treatment protocols. Progress in mapping of genes for mental retardation and epilepsy is discussed below.

1.8.1 Mental retardation

Mental retardation (MR) is a very common phenotype and is known to be caused by defects in a large number of genes (Antonarakis and Aelst, 1998). MR is described as significantly sub-average intellectual functioning and limitations in adaptive skills, with onset before 18 years of age (American Psychiatric Association, 1995). Approximately 2%-3% of the human population has either learning or behavior disabilities with an intelligence quotient (IQ) less than 70 and 0.3% of the individuals are severely handicapped (IQ<50) (McLaren and Bryson, 1987). The disorder has a substantial genetic component and there may be a genetic cause in approximately 50% of severely retarded patients (Hagberg and Kyllerman, 1983). In some cases mental retardation may be a part of a complex syndromes like Down syndrome, fragile X syndrome and ATR-X syndrome, metabolic disorder (example, phenylketonuria), or neuromuscular diseases like

Duchenne muscular dystrophy (Gecz and Mulley, 2000). In contrast, there is an increasing number of other conditions in which mental retardation is the only consistent clinical feature and these are called non-syndromic/non-specific MR (NSMR) (Toniolo and D'Adamo, 2000). Other than cognitive impairment, the NSMR affected patients do not have any distinctive clinical or biochemical features or brain development abnormalities (Chelly, 1999).

1.8.1.1 X-linked mental retardation

It is reported that 25%-50% of mental impairment is X-linked (Stevenson, 1996; Nigro et al., 2000). Because of the haploid status of most genes on the X-chromosome in males and the availability of families with mentally retarded males, mainly X-linked forms of mental retardation have been mapped (Chelly, 1999). The prevalence of X linked syndromal and nonsyndromal mental retardation cases is 1.66/1000 males (Glass 1991; Turner et al., 1996) and 0.9-1.4/1000 males (Kerr et al., 1991) respectively. The first gene, FMR1, for the most common form of syndromic mental retardation, the fragile X-A (FRAXA) syndrome, was identified in 1991 (Verkerk, 1991). FRAXA accounts for 15%-20% of all XLMR (Turner et al., 1996). Since then a number of genes involved in the syndromal forms of XLMR have been identified and cloned (reviewed by, Lubs et al., 1996; Chelly, 1999).

Approximately 65 NSMR loci corresponding to individual families are mapped along the X chromosome and can be grouped into 10-12 non-overlapping regions, suggesting the involvement of a minimum number of 10-12 X-linked genes in NSMR (Chelly, 1999). By 1999, seven genes related to non-specific X-linked mental retardation (MRX) were discovered: FMR2, GDI1, OPHN1, PAK3,

IL1RAPL, RPSK2 and TM4SF2 (Toniolo and D'Adamo, 2000; Gecz and Mulley, 2000). The absence or truncation of FMR2 protein results in FRAXE associated mental retardation (Gecz et al., 1996; Gecz, 2000a). FRAXE is a folate sensitive fragile site in Xq28 (Sutherland and Baker, 1992). Normal individuals show 6-25 copies of the repeat, while individuals expressing the fragile site have greater than 200 copies and their CpG island is methylated (Knight et al., 1993). Out of these seven MRX genes six (GDI1, OPHN1, PAK3, IL1RAPL, RPSK2 and TM4SF2) have been shown to participate in various stages of intracellular signaling (reviewed by Toniolo and D'Adamo, 2000; Gecz and Mulley, 2000).Recently in July 2000, a novel gene FMR3 associated with FRAXE was identified. The role of FMR3 in FRAXE MR is yet to be established (Gecz, 2000b).

The total number of genes involved in MRX is still a controversy. Early estimates of Morton et al., (1977, 1978) predicted the involvement of at least 17 X-linked genes causing non specific mental retardation on the basis of the mutation rate of 0.008 per gamete per generation or $<2.4 \times 10^{-5}$ per locus. Gedeon et al., (1996) predicted 8 genes for NSMR. However, in, 1999, this number was increased to 12 (Chelly, 1999). So far 8 genes for MRX have been identified. Additionally, DXS6673E (Van der Maarel et al., 1996) and GRIA3 (Gecz et al., 2000) have been identified through cloning of X-chromosome rearrangement breakpoints associated with MRX. However, these have not yet been found mutated in MRX families and also the involvement of autosomal genes has not been excluded in these cases. Based on the documented X-chromosome rearrangement cases related to MRX, it is now predicted that possibly the number of MRX genes may exceed 100 (Gecz and Mulley, 2000).

1.8.1.2 Autosomal mental retardation

Identification of families with mentally retarded males has led to the mapping and cloning of a number of X-linked genes for mental retardation. However, identification of equivalent genes on autosomes is more problematical. Autosomal recessive genes related to such disorders are unlikely to be identified except in specialized circumstances (example, defined populations with founder effects, families with inbreeding). Autosomal dominant forms may also be difficult to genetically map due to the likely presence of genetic heterogeneity if the phenotype is relatively non-specific. In addition, individual pedigrees are unlikely to be large when affected individuals have a relatively severe phenotype. Therefore, it is important to exploit any opportunity to clone autosomal MR genes. Positional cloning efforts based on the investigation of autosomal translocations, deletion mapping and candidate gene strategy have identified many genes involved in the syndromal forms of autosomal MR. As of December 30, 2000, there are nearly 949 entries related to MR in the OMIM database (URL W5, Table, 1.1), with approximately 79% related to syndromal autosomal recessive or dominant phenotypes. However, these genes may not be directly involved in the development of human intellectual and learning abilities as MR is a secondary feature in most of the syndromes (Chelly, 1999).

Mental retardation is a very common phenotype however, the etiologies of most non-syndromic cases of mental retardation are not known at this stage. Therefore, research in this field today focuses on the identification of patients with idiopathic MR having chromosomal rearrangements and subsequently identifying the genes disrupted at the breakpoints. There are several published reports in which individuals with balanced *de novo* translocations involving auotsomes have had

MR of varying degrees including the reports published by Hattori et al., (1985); St Clair et al., (1990); Reddy et al., (1999); Rosenberg et al., (1999). Hitherto, no autosomal gene for a non-specific form of mental retardation has yet been identified.

Another approach being practiced is to identify families associated with nonspecific mental retardation and map the gene by linkage analysis. Two such families associated with non-specific mental retardation were recently reported. Holinski et al (2000) studied a five-generation family that includes 10 individuals in generations IV and V who are affected with mild to moderate mental retardation and mild, nonspecific dysmorphic features. The disease in this family is inherited in a seemingly dominant fashion with reduced penetrance. This pedigree is unusual because of its size and the fact that individuals with the disease appear only in the last two generations. Standard clinical and laboratory screening protocols and extended cytogenetic analysis, including the use of high-resolution karyotyping and multiplex FISH (M-FISH) analyses, were unable to identify the cause of mental retardation in this family. Therefore, the group performed a whole genome scan by linkage analysis using microsatellite markers. The phenotype was linked to chromosome 16p13.3, and unexpectedly, a deletion of part of 16pter was demonstrated in patients, similar to the deletion observed in patients with ATR-16 syndrome, characterized by alphathalassemia and mental retardation. Subsequent FISH analysis demonstrated that the patients inherited a duplication of terminal 3q in addition to the deletion of 16p. FISH analysis of obligate carriers revealed that a balanced translocation between the terminal parts of 16p and 3q segregated in this family. The mental retardation in this family

appears to be ATR-16 syndrome due to an inherited cryptic (cytogenetically invisible) subtelomeric translocation, t(3;16)(q29;p13.3).

Recently in August 2000, another pedigree having non-syndromic mental retardation was identified with ten affected individuals (Higgins et al., 2000). This report suggests the localisation of a gene for recessive type of autosomal non-syndromic MR to 3p25-pter. To identify a MR disease locus in this family, a genome-wide search using genetic loci spaced at 10-20-cM was done. Genetic linkage between a MR disease locus and locus D3S3050 on chromosome 3p25-pter was established with a Zmax = 9.18 at $\theta = 0.00$. Fine mapping of this region narrowed the disease interval to 13.5-cM defined by recombinants at loci D3S3525 and D3S1304. Multipoint linkage analysis further refined the critical region to a 6.7-cM interval flanked by D3S3525 and D3S1560 DNA markers. However, the disease gene in this family has not yet been identified.

In conclusion, the existing literature suggests that there are many more X-linked and autosomal genes, yet unidentified, which might have a role to play in NSMR (Gecz and Mulley, 2000).

1.8.2 Genetic aspects of the epilepsies

Epilepsy comprises a heterogeneous group of seizure disorders, affecting 3% of the population during lifetime (Hauser et al., 1996; Sander et al., 2000). A seizure is a paroxysmal and transitory disturbance of brain function that develops suddenly and terminates spontaneously. It is physiologically characterised by abnormal, excessive and self-terminating discharges from neurons (Zara et al., 1995).

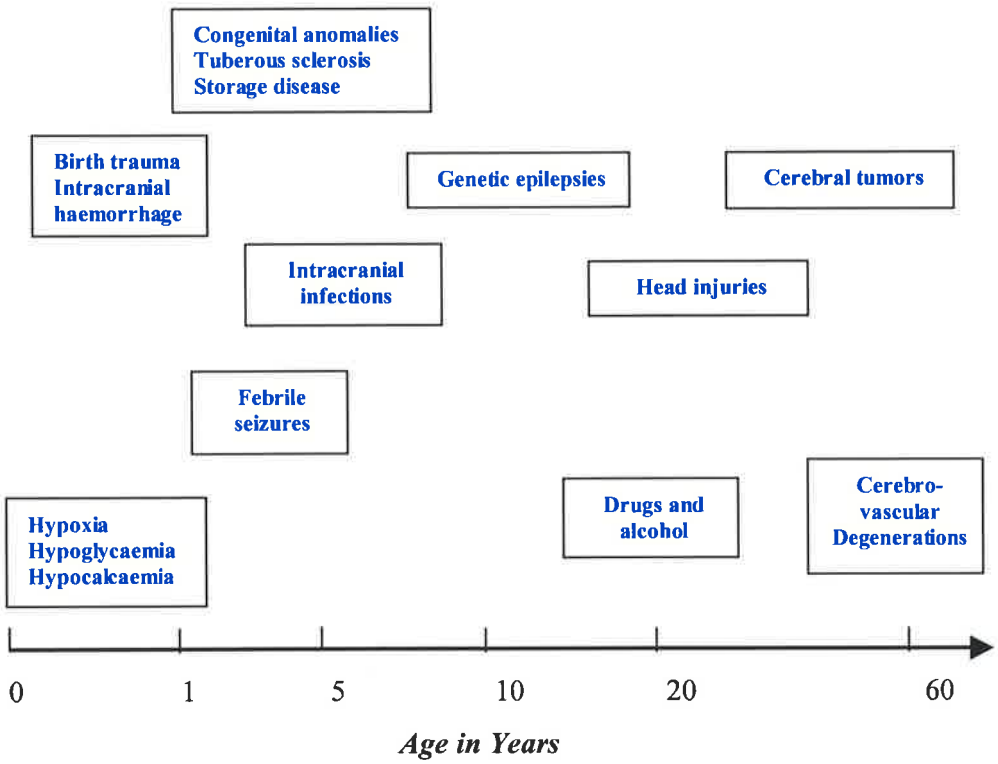


Figure 1.2 Causes of seizures and epilepsy by age
Reproduced from Epilepsy: A handbook, Buchanan N (1995)

Epilepsy is multifactorial and 60% of the patients may have no specific cause. It may be a symptom associated with any form of cerebral pathology which is in part age-related (Figure 1.2) (Buchanan, 1995). However, genetic aetiology may be present in up to 40% of patients and this proportion is even higher in epilepsy of childhood onset (Robinson and Gardiner, 2000). Currently, the focus of research is the identification of mutations causing epilepsies, the manner in which these epilepsy mutations are perpetuated, and the abnormal properties of the neuron glia system through which the mutations are expressed and result in a clinical phenotype (Delgado-Escueta et al., 1994).

1.8.2.1 Historical background and classification of epilepsy

Until recently, epilepsy was known as the sacred disease and accepted to be caused by supernatural powers. Epileptics were shunned and feared and saints were called to cleanse the sufferers, despite the fact that Hippocratic writers denied that epilepsy was sacred as long ago as the fourth century B.C. They believed that there was no difference between the sacred disease and any other and stated that epilepsy was inherited (Tempkin, 1971). This concept received a definitive stamp of proof when the first genetic defect for idiopathic epilepsy was identified in the α4 subunit (CHRNA4) of the neuronal nicotinic acetylcholine receptor, a ligand-gated ion channel, which is responsible for some families with autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE) (Steinlein et al., 1995). The gene for ADNFLE maps to chromosome 20q13.2-q13.3 (Phillips et al., 1995).

It is helpful to classify epilepsies according to their mode of inheritance and according to whether they are idiopathic (primary) or symptomatic types

(Robinson and Gardiner, 2000). According to the mode of inheritance, three major groups are recognized, Mendelian epilepsies, non-Mendelian or 'complex' epilepsies, and chromosomal disorders. In Mendelian epilepsies, a single major locus accounts for the segregation of the disease in the family. The second group is non-Mendelian or 'complex' epilepsies, where several gene loci together with environmental factors contribute to the disease phenotype. And in chromosomal disorders, a gross brain abnormality is present.

Over 160 Mendelian phenotypes are known where epilepsy is present as a component. Most are symptomatic and associated with major central nervous system abnormalities and are recognisable. However, there are a small but important number of 'idiopathic' Mendelian epilepsies (Table 1.2). Idiopathic epilepsies are mainly benign and represent about 30-40% of all epilepsies in childhood and about 20% in adult patients (Steinlein, 1998). This group includes epilepsies such as, benign familial neonatal or infantile convulsions, autosomal dominant nocturnal frontal lobe epilepsy and generalised epilepsy with febrile seizures (Robinson and Gardiner, 2000). Non-Mendelian 'complex' idiopathic epilepsies comprise seven clinically delineated syndromes out of which juvenile myoclonic epilepsy (JME), childhood absence epilepsy (CAE), juvenile absence epilepsy (JAE) and benign epilepsy with centrotemporal spikes (Table 1.2) represent the most common subtypes (Sander et al., 2000; Robinson and Gardiner, 2000).

Mendelian symptomatic epilepsies include progressive myoclonic epilepsies (PMEs). PMEs constitute a rare heterogeneous group of epilepsies characterized by the presence of myoclonus seizures and progressive neurological deterioration,

particularly dementia and ataxia (Serratosa et al., 1999). It accounts for 1% of all epilepsies occurring in childhood and adolescence (Robinson and Gardiner, 2000). PMEs (Table 1.2) can be caused by Unverricht-lundborg disease, juvenile type of ceroid lipofuscinoses (CLN3), Lafora disease and juvenile type of Gaucher's disease (Delgado-Escueta, 1994). Non-Mendelian 'complex' progressive myoclonic epilepsies include myoclonic epilepsy and ragged red fibers (MERRF) disorder. This is a mitochondrial disorder that is diagnosed histologically by the presence of ragged red fibers on skeletal muscle biopsy (Robinson and Gardiner, 2000).

1.8.2.2 Progress in mapping human epilepsy genes

(Table 1.2)

Significant progress has been made over the past few years in identification and mapping of the genes for epilepsy. Early efforts in epilepsy research identified the first epilepsy locus in a common benign idiopathic generalized epilepsy syndrome, juvenile myoclonic epilepsy (JME), in 1987 (Greenberg et al., 1988). Properdin factor, human leukocyte antigen (HLA) and DNA markers in the HLA-DQ region were genetically linked to JME and the locus was named as EJM1. JME was mapped to short arm of chromosome 6 (Greenberg et al., 1988; Delgado-Escueta, 1994). A gene for JME, CHRNA7, has now been cloned using candidate gene approach (Elmslie et al., 1997). CHRNA7 encodes the α7 subunit of the neuronal nicotinic acetylcholine receptor.

Progress in positional cloning of epilepsy genes in the monogenic forms of idiopathic epilepsies provide emerging evidence that these seizure traits arise from mutation in genes encoding ion channels (Ryan, 1999; Sander et al., 2000).

Table 1.2 Genetics of idiopathic and symptomatic epilepsies (modified from Robinson and Gardiner, 2000).

Epilepsy Syndrome	*Inheritance*	*Gene location*	*Gene*
Idiopathic epilepsies			
(Mendelian inheritance)			
Benign familial neonatal convulsions	AD	20q(EBN1)	KCNQ2
		8q (EBN2)	KCNQ3
Benign familial infantile convulsions	AD	19q	unknown
Autosomal dominant nocturnal frontal lobe epilepsy	AD	20q13.2	CHRNA4
		1p21.1-q21	CHRNB2
Generalized epilepsy with febrile Seizure plus	AD	19q13	SCN1B
		2	SCN1A
(Non-Mendelian inheritance)			
Juvenile myoclonic epilepsy	complex	15q14	?CHRNA7
		6p(EJM1)	unknown
Childhood absence epilepsy (and/or EEG trait)	complex	8q24	unknown
Juvenile absence epilepsy	complex	?21q22.1	?GRIK1
Benign epilepsy with centrotemporal spikes	complex	15q14	unknown
Symptomatic epilepsies			
(Mendelian inheritance)			
Unverricht-Lundborg disease	AR	21q22.3	EPM2
Neuronal ceroid lipofuscinoses			
Infantile	AR	1p32	CLN1
Classic late infantile	AR	11p15	CLN2
Finnish late infantile	AR	13q21-32	CLN5
Variant late infantile	AR	15q21-23	CLN6
Turkish variant late infantile	AR	unassigned	CLN7
Late infantile with GRODs	AR	1p32	CLN1
Juvenile	AR	16p12	CLN3
Variant juvenile with GRODs	AR	1p32	CLN1
Progressive epilepsy with MR	AR	8p23	CLN8
Kuf's disease	AR	unassigned	CLN4
Lafora disease	AR	6q24	EPM2A
Juvenile type of Gaucher's disease	AR	1q21-q31	Glucocere-brosidase gene
(Non-Mendelian inheritance)			
Myoclonic epilepsy and ragged red fibers (MERRF)	Mitochondri-al disorder		t RNA (Lys) gene

AD = autosomal dominant, AR = autosomal recessive, EEG = electroencephalogram, GROD = granular osmiophilic deposits, MR = mental retardation.

Mutations in two voltage–gated potassium channel genes (KCNQ2 and KCNQ3) cause benign familial neonatal convulsions (Singh et al., 1998; Charlier et al., 1998). Febrile seizures are the most common seizure disorders and affect 3% of all children under six years of age (Wallace et al., 1998; Scheffer et al., 2000). A mutation in the voltage-gated sodium channel β1 subunit gene (SCN1B) confers susceptibility to febrile seizures and generalized epilepsy ($GEFS^+$) in a large Australian family linked to 19q13.1 (Wallace et al., 1998). A new locus for generalized epilepsy with febrile seizure was recently mapped to chromosome 2 in large Australian family of northern Victoria (Lopes-Cendes et al., 2000).

There has been a controversy whether this new locus is associated with a newly described syndrome ($GEFS^+$) or with febrile seizures. Peiffer and colleagues argue that this is a third locus for common form of febrile seizures. The other two loci for the common form of febrile seizures are 8q13-21 (Wallace et al., 1996) and 19p13.3 (Johnson et al., 1998). However, Scheffer et al., (2000) on clinical grounds, believes that this new locus on chromosome 2 is associated with $GEFS^+$. Mutations in the neuronal sodium channel alpha-subunit gene, SCN1A, on chromosome 2q24, were recently described in two families with $GEFS^+$ (Escayg et al., 2000).

In at least two families autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE) arises from mutations in the gene encoding the α4 subunit of neuronal nicotinic acetylcholine receptor subunit (CHRNA4)(Steinlein et al., 1995). ADNFLE is a partial epilepsy causing frequent, violent, brief seizures at night, usually beginning in childhood (Steinlein et al., 1995). Recently Phillips et al., (2000), showed a *de novo* Ser252Leu CHRNA4 mutation in a 30 year old woman

diagnosed with sporadic nocturnal frontal lobe epilepsy. Phillips et al., (2000), have also identified recently a mutation in a Scottish family associated with ADNFLE. This mutation results in a V287M substitution within the M2 domain of a gene CHRNB2. CHRNB2 is a new gene for idiopathic epilepsy, the second acetylcholine receptor (β2) subunit implicated in ADNFLE. A point mutation in the human voltage-gated potassium channel gene Kv1.1 on chromosom12p13 is also reported to be associated with partial epilepsy (Zuberi et al., 1999). Association of common ion channel mutations to the monogenic epilepsies may shed some light on the understanding of the basic mechanisms of epileptogenesis.

Progressive myoclonic epilepsies (PMEs) are Mendelian or mitochondrial in origin (Serratosa et al., 1999). The gene defects for the most common forms of PME have been either identified or mapped to specific chromosome sites. Unverricht-Lundborg disease was shown to be caused by mutations in the gene (EPM2) that codes for cystatin B, an inhibitor of cysteine protease. The most common mutation in Unverricht-Lundborg disease is an expansion of dodecamer repeat in the non coding region upstream of the transcription start site of the cystatin B gene, making it the first human disease associated with instability of a dodecamer repeat (Serratosa et al., 1999). The neuronal ceroid lipofuscinoses are a group of at least ten neurodegenerative disorders characterised by the accumulation of autofluorescent lipopigment in neurons and other cell types (Robinson and Gardiner, 2000). Six genes for NCL have been mapped and four cloned, with at least two more to be identified (Bate and Gardiner, 1999; Robinson and Gardiner, 2000). A gene (EPM2) for progressive myoclonus epilepsy of the Lafora type (EPM2) has been mapped to chromosome 6q24, which encodes a protein, called tyrosine phosphatase (PTP). The known mutations in

this gene suggest that LD (Lafora disease) is caused by mutational inactivation of PTPase activity that may be important in the control of glycogen metabolism (Minassian et al., 1998; Serratosa et al., 1999). A juvenile type of Gaucher's disease is caused by the deficiency of glucocerebrosidase (Tsuji et al., 1987). A homozygous mutation in the glucocerebrosidase gene result in accumulation of neurotoxic glucocerebrosides leading to the disease phenotype. Myoclonus epilepsy with ragged-red fibres is caused by mutations in the mitochondrial gene that encodes for tRNA (Lys) (Serratosa et al., 1999).

Once a mutated epilepsy gene is identified the transgenic expression of these mutations can be studied using mouse models. Various mouse models for epilepsy are described in Noebels (1996) and Fletcher and Frankel (1999). This will allow a understanding of the critical steps involved in epileptogenesis, while the mutants themselves serve as reproducible biological test systems for therapeutic discovery (Noebels, 1996). With the increasing knowledge in the field of epilepsy, it is probable that in the next decade, the molecular basis for the familial complex epilepsies will be determined. It is also possible that new developments in DNA microarray technology will allow analysis of DNA from a buccal swab using a microarray chip designed to test for common mutations in, say, 50 ion channel genes. A precise molecular diagnosis will be available which will help to design and prescribe antiepileptic drug specifically for the particular electrophysiological dysfunction present (Robinson and Gardiner, 2000).

1.9 General conclusions

As discussed in sections 1.1 and 1.8, it is clear that 30,000 – 120,000 genes are expressed in man and over half may have a role to play in the brain structure and

function. Although many genes associated with syndromal forms of X linked and autosomal mental retardation and eight genes for X-linked non-syndromal mental retardation (sections, 1.8.1.1 and 1.8.1.2) have been isolated, the etiologies of most non-specific autosomal MR cases are unknown. There is only a single published report suggesting the localisation of a gene for autosomal NSMR to 3p25-pter (section, 1.8.1.2). Hitherto, there is no gene identified yet for the autosomal NSMR. Mental retardation in the second family discussed in section, 1.8.1.2 appears to be associated with ATR-16 syndrome. Mutations in some ion channel genes (refer, section 1.8.2.2) are now shown to be causes of epilepsy, but due to the genetic heterogeneity molecular mechanisms involved in the clinical phenotypes are not completely understood. Hence, to better understand the genetics underlying mental retardation and epileptic seizures, it is necessary to identify more disease genes associated with the functions of brain by exploiting any opportunity to clone such genes. Linkage analysis has been successful in mapping various gene loci for MR and epilepsy. But this technique requires a single large family to accurately map the disease-related gene, which is not always possible. Positional cloning strategies can be used to overcome this problem. Further, positional cloning can be greatly assisted by the identification of patients having chromosomal rearrangements associated with abnormal phenotypes. Two such patients were reported at the Department of Cytogenetics, WCH, Adelaide, with balanced *de novo* translocations associated with mental retardation and epilepsy. The translocation breakpoints in the two patients have been localized to chromosome 16 at 16p13.3 (Callen 1992, 1995). These two translocation cases provided a unique opportunity to clone the gene related to brain function in the two patients. Identification of the candidate gene related to

MR and/or epilepsy will further help in better understanding of the molecular basis of neuronal mechanisms underlying these disorders

1.10 The project

1.10.1 Prior knowledge

Two patients with mental retardation and epilepsy had associated *de novo* balanced translocations t(1;16) and t(14;16) respectively. Cell lines derived from these two patients had been used to generate mouse/human somatic cell hybrids. The hybrids containing the der(16) of the t(1;16) was designated as CY196 and that containing the der16 of the t(14;16) was designated as CY182. These two chromosome 16 breakpoints were included in the chromosome 16 somatic cell hybrid panel that was used to physically map chromosome 16 in detail with an average resolution of 1-Mb (Callen et al., 1992, 1995). The chromosome breakpoints in CY196 and CY182 resided in close vicinity on chromosome 16 at 16p13.3 (Figure 1.3). No DNA probes were mapped between these two breakpoints, which suggested they were less than one megabase apart.

1.10.2 Project hypothesis and aims

Hypothesis

It was hypothesized that mental retardation and epilepsy in two patients with *de novo* balanced translocations is due to a common gene defect at or near the 16p13.3 breakpoints of the t(1:16) and t(14;16).

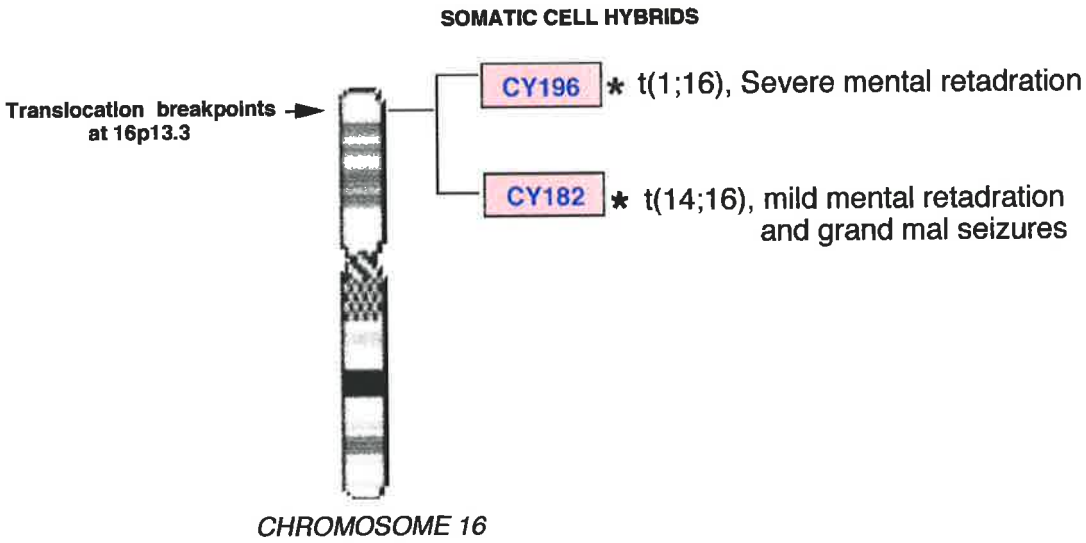


Figure 1.3 Physical representation of t(1;16) and t(14;16) translocations.

Project aims

The aim of the research presented in the thesis is to identify if any gene(s) was disrupted at the breakpoints of the t(1;16) and t(14;16) using positional cloning strategies.

Specific aims

A) Due to the lack of DNA markers within the hybrid interval defined by the CY196 and CY182 breakpoints, the initial aim of the project will be to generate a physical map of the breakpoint region by mapping STSs and ESTs in the vicinity of the breakpoints. Information regarding these markers will be obtained from the RH (radiation hybrid) map and CEPH (Centre d'Etude du Polymorphisme Humain) genetic map of chromosome 16.

B) These markers will be then used to identify YAC, BAC or PAC DNA clones spanning the breakpoints.

C) This will be followed by the identification of the candidate gene from the implicated region as discussed in section 1.5.

D) Finally, if time permits, unrelated patients with MR and/or epilepsy will be screened for mutations in the coding region of the gene to demonstrate if it is involved in sporadic disease.

Materials and Methods

Chapter 2

2.1 Introduction

Methods used in the present study are in common practice in the Department of Cytogenetics and Molecular Genetics, at the Women's and Children's Hospital, University of Adelaide, Australia. The methods described in this chapter are the basic molecular genetic methods used for DNA isolation, cloning of DNA, genomic DNA analysis by polymerase chain reaction and Southern hybridisation, expression analyses, mutation detection and DNA sequencing. The materials and methods used for a specific section of the project are presented at the beginning of the relevant thesis chapter. Restriction enzymes used in the project were obtained from commercial sources and used in accordance with the manufacturers' specifications. All solvents and chemicals were of analytical grade.

2.2 Materials

2.2.1 Enzymes and suppliers:

E. coli DNA polymerase (klenow fragment)	Amersham Australia Pty Ltd
Lysozyme	Boehringer Mannheim
Proteinase K	Sigma
RNase A	Boehringer Mannheim
Superscript (RNase reverse transcriptase)	Gibco BRL, Life Technologies
T4 DNA ligase	Boehringer Mannheim
Taq polymerase	Gibco BRL, Life Technologies or Boehringer Mannheim
X-Gal (5-bromo-4-chloro-3-inodolyl-β-D galactoside)	Progen

All restriction endonucleases were obtained from New England Biolabs (Beverly,

Massachusetts, USA), Progen (Brisbane, Queensland, Australia) and MBI Fermentas (Australia)

2.2.2 Electrophoresis reagent and suppliers

Agarose-nucleic acid grade	Pharmacia, Uppsala, Sweden
Agarose-low melting temperature	FMC, Rockland, Maine
Bromophenol blue	BDH chemicals, Dorset, England
Ethidium bromide	Boehringer Mannheim
TEMED (N,N,N',N'-tetramethyl-ethylenediamine)	BioRad
Xylene cyanol	Tokyo Kasei, Tokyo, Japan

2.2.3 Molecular weight markers

pUC 19/*Hpa*II	Bresatec, Adelaide, Australia
SPP1/*Eco*RI	Bresatec, Adelaide, Australia
DRIgest III	Pharmacia
Low range PFG marker	New England Biolabs
Yeast DNA-PFGE	Pharmacia

2.2.4 Radio-chemicals

α ^{32}P-dCTP, 3000 Ci/mmole	Radiochemical Centre, Amersham

2.2.5 Buffers and solutions

50X Denhardt's solution	5 g Ficoll
	5 g Polyvenylpyrrolidene

	5 g BSA Fraction
Formamide loading buffer	92.5% (v/v) Formamide
	20 mM EDTA
	0.1% (w/v) Xylene cyanol
	0.1% (w/v) Bromophenol blue
10X loading buffer	50% Glycerol
	1% (w/v) SDS
	100 mM EDTA
	0.1% (w/v) Xylene cyanol
	0.1% (w/v) Bromophenol blue
2X PCR mix	33 mM Ammonium sulphate
	133 mM Tris-HCl
	2% (v/v) β Mercaptoethanol
	13 mM EDTA
	0.34 mg/ml BSA
	20% (v/v) DMSO
	3 mM dATP, dGTP, dTTP, dCTP
20X SSC	3 M NaCl
	0.3 M tri-sodium citrate (pH 7.0)
20X SSPE	3.6 M NaCl
	0.2 M $NaH_2PO_4.2H_2O$
	0.02 M EDTA
TBE	89 mM Tris-base
	89 mM Boric acid
	2.5 mM EDTA (pH 8.3)
TE	10 mM Tris-HCl (pH 7.5)

0.1mM EDTA

2.2.6 *Bacterial and yeast media*

2.2.6.1 Liquid media

All liquid media were prepared with millipore water and autoclave sterilised. The compositions of the various medias are as follows:

AHC medium (for *S.cerevisiae*)	0.67% (w/v) yeast nitrogen base without amino acids 1% (w/v) casein hydrolysate-acid 0.006% Adenine 2% Glucose Tetracycline (50 μg/ml)
L. (Lauria) broth (for *E. coli*)	1% (w/v) Bacto-triptone 0.5% (w/v) Bacto-yeast extract 1% (w/v) NaCl (pH to 7.5 with NaOH)
Glycerol broth	Lauria broth 15% Glycerol

2.2.6.2 Solid media

AHC agar	AHC media (broth) Bacto-agar (1% w/v)
L. agar	L. broth 1% (w/v) Bacto-agar
L. ampicillin agar	L. broth 1% (w/v) Bacto-agar

	Ampicillin (100 µg/ml)
L. kanamycin agar	L. broth
	1% (w/v) Bacto-agar
	Kanamycin (50 µg/ml)
L. chloramphenicol agar	L. broth
	1% (w/v) Bacto-agar
	Chloramphenicol (30 µg/ml)
L. tetracycline agar	L. broth
	1% (w/v) Bacto-agar
	Tetracycline (20 µg/ml)

2.2.7 Antibiotics

The antibiotics used in this project are listed below with the names of the suppliers

Ampicillin	Sigma
Chloramphenicol	Sigma
Kanamycin	Progen
Tetracycline	Sigma

2.2.8 Bacterial strains (E. coli)

XL1-Blue	Genotype: rec A1, end A1, gyr A96, thi -1, hsd R17, sup E44, re1 A1, lac [F^1 pro AB, lac 1q zamI5, tn10 (tetr)] Supplier: Stratagene, California, USA

2.2.9 Vectors

pUC 19	New England Biolabs
pGEM-T	Promega

2.2.10 Miscellaneous materials

DNA isolation kits	Qiagen, Hilden Germany
DNA sequencing kits	ABI-Perkin Elmer, Warrington, UK
HyBond N^+ Nylon membrane	Amersham, USA
Multiple Tissue Northern blots	Clontech, USA
PAC/BAC Human library sets	Genome Systems. St Louis, Missouri; PAC/BAC Resources, Buffalo, NY
PGEM-T cloning kit	Promega, USA
Qiaquick column purification kit	Qiagen, Hilden, Germany
5' RACE kit	Gibco BRL, Life Technologies, Gaithersburg, USA
RNeasy Midi kit	Qiagen, Hilden, Germany
RT PCR reagents kit	Gibco BRL, Life Technologies, Gaithersburg, USA
TRIzol reagent	Gibco BRL, Life Technologies, Gaithersburg, USA
X-ray film	Kodak or Dupont

2.2.11 Miscellaneous fine chemicals

Adenine	Sigma
Ammonium sulphate	Ajax Chemicals, Australia
Bacto-agar	Sigma Chemical Co., USA

Bacto-triptone	Difco Laboratories, USA
Bacto-yeast extract	Gibco BRL
βMe (β-mercaptoethanol)	BDH chemicals Ltd., Poole, England
Boric acid	Ajax
BSA (bovine serum albumin)	New England Biolabs
Casein hydrolysate-acid	ICN Biomedicals, Ohio
Chloroform	BDH chemicals
Deoxynucleotide triphosphates and Dideoxynucleotide triphosphates	Boehringer Mannheim
Dextran sulphate	Pharmacia
Diethyl pyrocarbonate (DEPC)	BDH
Dimethylsulphoxide (DMSO)	Sigma
Ethanol (99.5% v/v)	BDH
EDTA (ethylenediaminetetracetic acid; $Na_2EDTA.2H_2O$)	Ajax
Ficoll	Sigma
Formamide	Fluka Chemika
Glucose	Ajax
Glycerol	Ajax
Human placental DNA	Sigma
Isoamy alcohol (IAA)	Ajax
Isopropanol	Ajax
Isopropylthio-β-galactoside (IPTG)	Boehringer Mannheim
Magnesium chloride ($MgCl_2.6H_2O$)	Ajax
Mixed bed resin (20-50 mesh)	BioRad, USA
Paraffin oil	Ajax

Phenol	Wako, Japan
Polyvenylpyrrolidene	Sigma
Salmon sperm DNA	Calbiochem
Sodium chloride (NaCl)	Ajax
tri-sodium citrate	Ajax
Sodium dihydrogen orthophosphate ($NaH_2PO_4.2H_2O$)	Ajax
Sodiumdodecyl sulphate (SDS)	Sigma
Sodium hydroxide (NaOH)	Ajax
Tris-base	Boehringer Mannheim
Tris-HCl	Boehringer Mannheim
Yeast nitrogen base	Difco laboratories

2.3 Methods

Most of the procedures described in this section are based on those presented in Molecular Cloning: A Laboratory Manual (Sambrook et al., 1989) unless otherwise specified.

2.3.1 DNA isolation

2.3.1.1 Plasmid DNA isolation

A Plasmid DNA was extracted from a 10 ml overnight culture using the Rapid Pure Miniprep (RPM^{TM}) kit (Qiagen, Hilden Germany). Manufacturer's protocol recommended 1.5 ml of the starting bacterial culture. This was modified by using 10 ml of the starting bacterial culture and doubling the value of reagents recommended for isolating DNA from 1.5 ml of the bacterial culture.

I) The culture from a bacterial stab/glycerol stock was streaked on a L. ampicillin.

agar (section, 2.2.6.2) to obtain a single isolated colony. This was inoculated in 10 ml L. broth (section, 2.2.6.1) containing 100 µg/ml of ampicillin. The culture was incubated overnight in a 37°C shaker incubator. Next day the 10 ml overnight culture was centrifuged at 4,000 g for 15 minutes at 4°C. The supernatant was discarded. The pellet was resuspended in 100 µl of kit-supplied pre-lysis buffer and transferred to an eppendorf tube. 200 µl of kit-supplied alkaline lysis solution was added and the contents then mixed until clear and viscous.

II) To the cell lysate, 150 µl of kit-supplied neutralising solution was added and the solution centrifuged at maximum speed (13,200 rpm) in a microfuge for 10 minutes. The supernatant was then transferred to a kit-supplied spin filter and 250 µl of glassmilk added. The solution was mixed and centrifuged at maximum speed in a microfuge for 1 minute.

III) The liquid was removed from the collection vial and 350 µl of kit-supplied wash solution was added to the spin filter and centrifuged at maximum speed in a microfuge for 1 minute. The spin was repeated and the spin filter transferred to a clean collection vial.

IV) The spin filter was allowed to dry for 10 minutes before adding 50 µl of sterile water. To resuspend the glassmilk, the spin filter was vortexed and centrifuged at maximum speed in a microfuge for 1 minute to collect the plasmid DNA.

2.3.1.2 Cosmid DNA isolation

Cosmid DNA was isolated from a 200 ml overnight culture, using a Qiagen Tip-20 column kit (Qiagen Plasmid Mini Handbook, 1995) as per the manufacturer's protocol.

I) The culture from a bacterial stab/glycerol stock was streaked on a L.

kanamycin (section, 2.2.6.2) agar plate to obtain single isolated colony. This colony was inoculated in 200 ml L. broth containing 50 µg/ml of kanamycin. The culture was incubated overnight in a 37°C shaker incubator. The overnight culture was then aliquotted into four 50 ml falcon tubes, centrifuged for 15 minutes at 4°C and the supernatant discarded. The four pellets were each resuspended in 500 µl of kit-supplied P1 resuspension buffer, combined and equal (350 µl) aliquots each placed into six eppendorf tubes.

II) To these tubes 350 µl of kit-supplied P2 lysis buffer was added, and the sample mixed, before adding 350 µl of kit-supplied P3 neutralisation buffer. The contents were mixed gently, tubes placed on ice for 10 minutes, and then centrifuged at 13,200 rpm in a microfuge for 30 minutes.

III) Each Qiagen tip-20 column was equilibrated with 1 ml of kit-supplied QBT buffer. The supernatant from step II was passed through the column, which was then washed with 4 ml of kit-supplied wash buffer. Subsequently, the column was placed over a fresh eppendorf and 800 µl of the kit-supplied elution buffer added.

IV) To elute DNA, 560 µl of isopropanol was added, the sample mixed and centrifuged at maximum speed in a microfuge for 30 minutes. The supernatant was carefully removed, replaced with 70% ethanol and centrifuged for a further 10 minutes to wash the pellet. The 70% ethanol was removed and the DNA pellet allowed to air dry by leaving the pellet at room temperature for one hour. The cosmid DNA was resuspended in 20µl of sterile water.

2.3.1.3 PAC and BAC DNA isolation

The PAC and BAC DNA was isolated using Qiagen Tip-100 column kit (Qiagen, Hilden Germany). The PAC and BAC DNA received as bacterial stabs were initially streaked on L. kanamycin and L. chloramphenicol agar (section, 2.2.6.2)

plates to obtain isolated colonies. For PAC DNA isolation, a single PAC colony was grown overnight at 37°C in 10 ml of L.broth containing 50 µg/ml of kanamycin. Next day, 200 ml of the L.broth plus 50 µg/ml of kanamycin was seeded with 6.7 ml of the 10 ml overnight culture and grown for 1.5 hours in a 37°C shaker incubator. After 1.5 hours incubation 100 µl of 1 M IPTG was added to each of 200 ml culture and cells were grown further for 5 hours at 37°C. In the case of BAC DNA isolation 200 ml of L broth containing 30 µg/ml of chloramphenicol was seeded with a single BAC colony and the culture was grown overnight in a 37°C shaker incubator. The DNA was then isolated using following steps.

I) Cell disruption:

200 ml of bacterial culture was centrifuged in a 250 ml Beckman bottle at 4,000 g in a Beckman Ultracentrifuge. The bacterial pellets were resuspended in 4 ml of kit supplied P1 buffer.

II) Cell lysis:

The resuspended pellets were transferred to 50 ml Oakridge tubes. To the resuspended pellet P2 lysis buffer (Table 2.1) was added. The suspension was mixed gently by inverting the tube six to eight times. The mixture was incubated at room temperature for no longer than five minutes.

III) Neutralisation:

The kit supplied neutralisation buffer P3 (Table 2.1) was added to the lysed cells and the suspension was mixed gently until a homogeneous suspension was formed. The suspension was incubated on ice for ten minutes and centrifuged at high speed for 18,000 g for 25 minutes. Supernatant from this step was transferred to a clean 50 ml Oakridge tube and centrifuged for a further ten minutes at 18,000

g to obtain a clear supernatant. If supernatant from this step is not clear, it will clog the cartridge, thereby preventing the adsorption of nucleic acids.

The supernatant of step III was removed carefully from the white precipitate and loaded on Qiagen tip-100 cartridge pre-equilibrated with buffer QBT. The cartridge was washed twice with wash buffer. The DNA was then eluted with pre-warmed (65°C) elution buffer. A second elution step was recommended to increase the DNA yield by 30%. The volumes of different buffers used for PAC and BAC DNA isolation are given below in Table 2.1. DNA was precipitated using 0.7 volume of isopropanol (Table 2.1). The DNA was subsequently treated as for cosmid DNA (section, 2.3.1.2). The dried pellets were resuspended in 5µl of sterile water.

Table 2.1 Working amounts of kit supplied buffers for PAC/BAC DNA isolation.

	PAC	BAC
Antibiotics	Kanamycin (50 µg/ml)	Chloramphenicol (30 µg/ml)
Buffer P1	4 ml	10 ml
Buffer P2	4 ml	10 ml
Buffer P3	4 ml	10 ml
Buffer QBT	4 ml	4 ml
Wash buffer	2X5 ml	2X5 ml
Elution buffer (pre-warned at 65°C)	2X2 ml	2X2 ml
Isopropanol	560 µl/800 µl of the elutent	560 µl/800 µl of the elutent

The above volumes were used when isolating DNA from 200 ml bacterial cultures.

In some cases the Nucleobond kit-tip 100 columns (Nucleobond, Duren) were used for the isolation of PAC/BAC DNA. The methods were the same as described for the Qiagen Tip-100 columns. The volumes of kit-supplied buffers used were as per the manufacturer's protocol.

2.3.1.4 Preparation of bacterial glycerol stocks

All bacterial clones were maintained as glycerol stocks that were prepared from a 10 ml of an overnight bacterial culture, which was being used for DNA isolation (section, 2.3.1) purposes. The culture was spun at 4,000 rpm and the pellet resuspended in 1 ml of L. glycerol broth (section, 2.2.6.1). These stocks were kept at –70°C until reused for DNA isolation.

2.3.1.5 Preparation of sheared human placental DNA and salmon sperm DNA

The commercially obtained human placental DNA (Sigma) and salmon sperm DNA (Calbiochem) were reconstituted in sterile TE to a concentration of 5 mg/ml and 20 mg/ml respectively. Aliquots of 1 ml were then transferred to a screw cap tubes. The tubes were maintained at 100°C for 20 hours. This was done to denature the DNA to approximately 700 bp fragments. The denatured DNA was then visualised on 1.5 % agarose gel. The concentrations of DNA samples were estimated by spectrophotometry (refer, 2.3.1.6) and adjusted to 5mg/ml in case of human placental DNA and to 20 mg/ml in case of salmon sperm DNA.

2.3.1.6 DNA quantitation

To determine the concentration of DNA, the optical density (OD) of the diluted DNA (1:250 in sterile distilled water) was measured on a spectrophotometer

(Ultrospec 3000, UV/Visible Spectrophotometer, Pharmacia, Biotech) at a wavelength of 260 nm. The absorbance was multiplied by the dilution factor and a conversion factor for particular type of DNA being quantitated. For a double stranded DNA, the conversion factor is 50 and for a single stranded DNA and RNA the conversion factor is 40. Using Beer's law and the known conversion factors, the concentration in each case was calculated using the formula:

DNA μg/μl = absorbance at OD_{260} X dilution factor (250) X (conversion factor)/1000

2.3.2 General methods for purification of DNA

2.3.2.1 Genomic DNA purification

One to two μg of PAC or BAC genomic DNA was digested with the appropriate restriction enzymes in a total volume of 50 μl. The entire reaction was purified using standard phenol/chloroform extraction (2.3.2.1.1) to remove enzyme and buffers.

2.3.2.1.1 Phenol chloroform extraction of DNA

(Modification of Moore, 1993)

An equal volume of TE saturated phenol was added to the digested DNA reaction (50 μl) and mixed by vortexing for one minute. The mixture was then centrifuged at maximum speed (13,200 rpm) in an eppendorf microcentrifuge for ten minutes. The upper phase was then transferred to a clean tube leaving a white phase of denatured protein and the lower organic phase. To the aqueous phase was added an equal volume of chloroform: isoamyl alcohol (24:1). Vortexing and centrifugation steps were repeated. Aqueous phase was again transferred to a clean tube. 3 M sodium acetate was added to the collected aqueous phase in the

ratio of 1:9 (volume:volume). To this was added 100% ethanol, which was 2.5 times of the total volume of the aqueous phase and added sodium acetate. The sample was then placed on ice for 30 minutes to allow the complete precipitation of DNA. The tubes were centrifuged at a maximum speed (13,200 rpm) in a microfuge for 30 minutes. The supernatant was removed, the pellet was washed with 70% ethanol and resuspended in 20 µl sterile water.

2.3.2.2 Cleaning of PCR and RT-PCR products

To generate PCR amplified DNA, a PCR reaction of 50-100 µl was used. A 5 µl sample of the PCR product was electrophoresed on agarose gel stained with ethidium bromide and visualised under UV light, to ensure that the band of correct size had been amplified. The remainder of the PCR reaction was cleaned by either using Qiaquick PCR purification kit (Qiagen, Hilden Germany)(section, 2.3.2.2.1) or any of the agarose gel DNA extraction methods described in section 2.3.2.2.2.

2.3.2.2.1 Qiaquick PCR purification method

The PCR reaction was mixed with five volumes of the Qiaquick buffer PB. The sample was loaded to a kit-supplied spin column placed in a 2 ml collection tube. Tubes were centrifuged for one minute and the solution in the collection tube was discarded. To wash the DNA, 750 µl of the Qiaquick buffer PE was added to the spin column and centrifuged for one minute. The flow through was removed and the column recentrifuged to remove any remaining PE buffer. The spin column was then transferred to a clean eppendorf tube and allowed to air dry for five minutes. Once the column was thoroughly dried 50 µl of sterile water was added to elute the DNA and left at room temperature for five minutes. The column was centrifuged for one minute and the DNA collected in a fresh tube.

2.3.2.2.2 Agarose gel DNA extraction methods

The Specific DNA bands of interest were excised from the agarose gels. The excised bands were trimmed to remove the excess of agarose, chopped with a sharp blade into very fine gel slices and DNA was eluted using following methods.

Prep-A-gene DNA Purification

Perp-A-gene matrix kit (BioRad) was generally used to clean the digested plasmid inserts excised from the gel. The concentration of the DNA in the band o the gel was estimated by comparison with bands generated from DNA of known concentrations. The volume of the excised gel slice was estimated by placing a similar tube next to it and adding liquid from a pipette into the empty tube until the volume is equivalent to that of the gel slice. Based on the volume of the gel slice plus the amount of Prep-A-Gene matrix required for DNA binding (for 1 μg of DNA or less a minimum of 5 μl Prep-A-Gene matrix is required), 3 volumes of kit-supplied binding buffer was added. The reaction was gently mixed and placed in water bath at 55°C for ten minutes to assist in the dissolving of the agarose. The predetermined amount of the Prep-A-Gene matrix was then added and the tube mixed for ten minutes at room temperature in a laboratory suspension mixer (Clements, Australia). The DNA bounded to the matrix was then pelleted by centrifuging for thirty seconds at a maximum speed in a micro-centrifuge. The supernatant was discarded and the DNA pellet rinsed by resuspending it gently in an amount of the binding buffer equivalent to 25 times the amount of added matrix, using brief vortexing. The centrifugation and the rinse step was then repeated. The Prep-A-Gene pellet was washed three times with 25X-matrix volume of wash buffer. After final washing step the pellet was dried for 5 minutes

in SpeedVac Concentrator. To elute the DNA 7.5 µl of sterile water was added to the pellet. The reaction was incubated at 42°C for 5 minutes. The pellet was briefly vortexed and centrifuged for 30 seconds. The supernatant was collected in a fresh tube and the DNA pellet was subjected to a second elution step for maximum recovery of DNA. The supernatant containing DNA was finally centrifuged again for 30 seconds to remove any traces of matrix.

Agarose gel DNA extraction kit

Agarose gel DNA extraction kit supplied by Boehringer Mannheim was used to purify PCR products ranging from 500bp-1kb in size. The DNA of interest was extracted (refer section, 2.3.2.2.2) in a preweighed screwcap tube. 300 µl of the agarose solubilisation buffer was added per 100 mg of the agarose gel. 10 µl of the silica suspension was added to the above tube. The amount of the silica suspension was increased by 4 µl/µg in the reaction when the amount of DNA was greater than 2.5 µg. The mixture was then incubated at 60°C for ten minutes and vortexed every 2-3 minutes. The reaction was centrifuged in a tabletop centrifuge for thirty seconds at 13,200 rpm. The supernatant was discarded and the matrix containing the DNA was resuspended in 500 µl nucleic acid binding buffer using a vortex mixer. The tube was centrifuged and the supernatant discarded as before. The pellet was then washed twice with 500 µl washing buffer. The supernatant from the washing step was discarded and the residual liquid removed with a pipette. The pellet was dried at room temperature for 15 minutes by inverting the tube on an adsorbent tissue. The DNA was eluted using 2X 25 µl of sterile water. Each elution step consisted of brief vortexing, incubation of ten minutes at 60°C followed by centrifugation at maximum speed

(13,200 rpm) for thirty seconds. The solution containing the DNA was transferred to a fresh tube and the tube was centrifuged for thirty seconds to remove any residual matrix.

DNA purification using Costar spin X centrifuge tube filter

(Schwarz and Whitton, 1992)

Costar spin X centrifuge columns (Catalogue number, 8160) were particularly useful for purification of smaller DNA bands of about 100 bp in size. The specific band of interest was excised from the gel (refer section, 2.3.2.2.2) and placed in the filter cup, mixed with 100-150 µl of distilled water or TE and reduced to fine slurry with the help of a thin steel rod. The spin column was then spun in a microcentrifuge at full speed for 5 minutes. The DNA was expelled into the collection tube while almost all of the gel material was retained in the filter cup. Generally 80% of the DNA was recovered using this method.

2.3.3 Subcloning of human DNA fragments

(Modification of Maniatis, 1982)

2.3.3.1 Preparation of plasmid vector and human DNA inserts

The pGEM-T vector used for cloning experiments was purchased from Promega. Human DNA inserts and PCR products to be ligated were purified using methods described in section 2.3.2.

2.3.3.2 Ligation reaction

For each DNA to be cloned (genomic or PCR product) a range of vector: insert ratios were used from 1:1 to 1:3 to maximise the intermolecular ligation ratio

compared with the intramolecular ligation. Typically, the 10 µl ligation reaction contained 50 ng of pGEM-T vector (Promega), approximately 50-300 ng of DNA insert, 1 µl (3 Weiss units) of T4 DNA ligase and 1µl of T4 DNA ligase 10X buffer. Ligation reactions were incubated overnight at 14°C.

2.3.3.3 Transformation

The plasmid vector containing the DNA to be cloned was transformed into *E. coli* XL1-Blue competent cells.

2.3.3.3.1 Preparation of competent cells

Method used from Chung et al., (1989)

XL1-Blue cells were streaked for single colonies on a L. tetracycline plate (section, 2.2.6.2) from a 5 µl aliquot of a glycerol stock. A single XL1-Blue colony was inoculated into the L. broth (section, 2.2.6.1) containing 20 µg/ml of tetracycline and incubated overnight in a 37°C shaker incubator. Stationary phase XL1-Blue cells from an overnight culture were diluted 1:50 (v/v) with fresh media (400 µl in 20 ml L. broth + 20 µl of tetracycline) and incubated at 37°C with shaking until the OD_{600} was approximately 0.3 (2-3 hrs). The cells were then centrifuged in a Jouan centrifuge at 3,000 rpm for 15 minutes at 4°C and the supernatant discarded. The cells were then gently resuspended in 1/10 volume (2 ml) of fresh TSS (1 g PEG 3350, 0.5 ml DMSO, 0.5 ml 1 M $MgCl_2$ and the volume was adjusted to 10 ml with L. broth) and placed on ice for at least 10 minutes prior to transformation. The prepared competent cells were aliquotted into ice-cold eppendorf tubes (300 µl/tube), and could be frozen in liquid nitrogen before use. The frozen cells were stored at –70°C until required.

2.3.3.3.2 Transformation using the 'heat shock' method

Method used from Chung et al., (1989)

A total of 5 µl ligation reaction was added to 100 µl of competent cells (section, 2.3.3.3.1), mixed gently and then placed on ice for 15-20 minutes. The cells were then heat shocked by exposure to 42°C for 1 minute and immediately returned to ice for 1 minute. 500 µl of L. broth was then added to these cells and incubated for 45 minutes at 37°C in a shaker incubator. Meanwhile, to each L. broth plate 40 µl of 20 mg/ml X-Gal and 20 µl of 200 mM IPTG were added and spread with a sterile glass spreader. After the 45 minutes incubation the entire transformation reactions (600 µl) were plated on the above-prepared L. broth plates. The plates were incubated overnight at 37°C. Recombinant plasmids were detected as white colonies with blue colonies containing only recircularised plasmid vector.

2.3.3.4 Colony PCR

The white colonies containing recombinant plasmids were picked and streaked onto a fresh L. agar plate containing the appropriate antibiotic (master plate) using a sterile pipette tips. The plate was incubated overnight at 37°C. A single white colony was picked up from the master plate and added to a PCR tube containing 35 µl of sterile water. A PCR mix was prepared containing 5 µl of 10X PCR buffer, 2 mM dNTPs, 150 ng/µl of each of pUC forward and reverse primers and 2.5 µl of 1/10 diluted Taq polymerase enzyme. The bacterial colony was resuspended in water by brief vortexing. 14.5 µl of the master mix was added to each tube. The PCR was performed in a Corbett PCR machine (FTS-960 Thermal sequencer, Corbett Research). A negative control (no DNA) was included in each experiment. The PCR cycle conditions used were denaturing at 94°C for 30 seconds, annealing at 60°C for 30 seconds and extension at 72°C for 1 minute.

This was repeated for 30 PCR cycles, followed by a final extension of the PCR products at 72^0C for 5 minutes. The PCR products were visualised on agarose gel. If an intense band of the expected size was detected, the plasmid was assumed to contain an insert. The DNA sequence (sections, 2.3.12.1 and 2.3.12.2) of the inserts was then obtained to confirm the integrity of the insert.

2.3.4 Restriction endonuclease digestion, gelelctrophoresis and Southern blot analysis of DNA

2.3.4.1 Restriction endonuclease digestion of cosmid, PAC/BAC DNA

Restriction endonuclease digestion of DNA was carried out in enzyme compatible buffers purchased from the New England Biolabs, Progen or MBI commercial buffer systems. All digests were performed at 37°C unless otherwise specified by the manufacturer. In general, 10 units of enzyme were added for 500 ng of cosmid DNA and 1.5 microgram of PAC/BAC DNA to be digested in the presence of 10X enzyme specific buffer (4 µl) in a final volume of 40 µl reaction. For double digests, a restriction enzyme buffer was used which was compatible for both enzymes (New England Biolabs catalogue, 1995). BSA (New England Biolabs catalogue number, B90015) at a final concentration of 100 µg/ml was added as per the manufacturer's recommendations. The reaction mix was incubated overnight at 37°C. The reactions were terminated by keeping the tubes at 65°C to inactivate the restriction enzymes or by adding 0.1X volume of 10X loading buffer.

2.3.4.2 Restriction endonuclease digestion of plasmid DNA

The plasmid DNA was generally digested for the isolation of cDNA inserts. 20 units of restriction enzyme were added for 5 µg of DNA to be digested in the

presence of 10 X specific restriction enzyme buffer in a 50 µl final reaction volume. The cDNA insert specific bands were excised form the agarose gel and cleaned as described in section 2.3.2.2.2.

2.3.4.3 Restriction endonuclease digestion of genomic DNA

10 µg of the genomic DNA or somatic cell hybrid DNA was digested in the presence of 20 units of restriction enzyme (*Pst*I or *Hin*dIII) and 10 X restriction enzyme specific buffer in a 50 µl final reaction volume. The separation of restriction enzyme digested genomic DNA or somatic cell hybrid DNA was achieved using 0.8% agarose gel (section, 2.3.4.4.1).

2.3.4.4 Gel electrophoresis

Restriction endonuclease digested DNA samples for Southern blot analysis were either resolved by agarose gel electrophoresis or pulsed-field gel electrophoresis (PFGE) depending on the size of digested DNA.

2.3.4.4.1 Agarose gel electrophoresis

Resolution of DNA in the range less than 30-40 kb was achieved by agarose gel electrophoresis (Sambrook et al., 1989). The separation of restriction enzyme digested plasmid DNA and cosmid DNA was performed in 0.8% agarose gel made in 1X TBE. Gels were run in BRL (BioRad) horizontal tanks. The smaller size PCR products 500 bp or less were eletrophoresed using 1-2% agarose gel in 1X TBE depending on the size of the expected product to be analysed, and run at 100 volts in a Bio-Rad Mini-Sub™ DNA cell. The separation of DNA fragments for restriction digested genomic DNA or somatic cell hybrid DNA was achieved using 0.8% agarose gel in 1X TBE and the gels were run overnight at 18 mA per

gel. Agarose loading buffer was added to each DNA sample (1X final concentration) prior to loading. In all cases, the gels contained ethidium bromide (0.0025 mg/ml) and DNA was visualised under UV light (Gel Doc 1000, BioRad) after the gel run.

2.3.4.4.2 Markers used in agarose gel electrophoresis

The choice of DNA marker was made depending on the expected size of the DNA fragment to be electrophoresed. Generally pUC19/*Hpa*II (size range 60 to 500 bp), SPP1/*Eco*RI (size range 0.5 to 8.5 kb), or Drigest (size range 72bp to 23 kb) were used as molecular weight markers on agarose gels.

2.3.4.4.3 Pulsed-field gel electrophoresis (PFGE)

Resolution of endonuclease digested DNA of YACs, BACs and PACs DNA clones was achieved by PFGE (Schwartz et al., 1982; Chu et al., 1986). A Bio-Rad CHEF Mapper (CA, USA) was used for PFGE. The digested DNA was electrophoresed in 1% agarose gel using 0.5X TBE for 28-30 hrs at 14°C. The gel was cast in PFGE support tray (14 cm X 12.7 cm). The operation of the PFGE apparatus was done as per the manufacturer's instructions.

2.3.4.4.4 PFGE markers

A Low range PFG marker (New England Biolabs), λ*Hin*dIII, and Drigest, from Pharmacia, were generally used for resolution of bands ranging from 20 kb – 1 Mb.

2.3.4.5 Labelling of DNA probes

(Sambrook et al., 1989)

DNA was labelled using the Amersham Megaprime^{TM} DNA labelling systems kit (Catalogue number, RPN, 1604) as per the manufacturer's instructions. A total of 50-100 ng/µl of DNA was used with the reaction volume increased to 34 µl with sterile water. The reaction was mixed and incubated at 100°C for 5-10 minutes. This was followed by the addition of 5 µl of 10X dNTP (dATP, dGTP, and dTTP), 5µl of 10X buffer containing random hexamers, 5 µl of α^{32}P-dCTP and 1 µl (5 units/µl) klenow DNA polymerase. During the process of adding these reagents the reaction was maintained at 37^{0}C. After adding these reagents the reaction was further incubated at 37^{0}C for 30 minutes. The reactions were then stopped by adding 1.5 µl of 0.5 M EDTA. Unincorporated radionucleotides were removed from the labelled probes using Qiaquick PCR column purification methods as described in section 2.3.2.2.1.

2.3.4.6 Pre-reassociation of repetitive DNA

Radiolabelled DNA probes containing repetitive sequences were pre-reassociated prior to DNA hybridisation (Sealy et al., 1985). The human placental DNA (section, 2.3.1.5) was denatured for 10 minutes and cooled on ice for 1 minute before adding to the probe. To the α^{32}P-dCTP labelled DNA was added 100 µl of (5 mg/ml or 2000 fold excess) denatured human placental DNA and 50 µl of 20X SSC. The samples were denatured in a 100°C-heating block for 10 minutes, cooled on ice for 1 minute and then incubated at 65^{0}C for 1 hr. After incubation, the sample was added to the hybridisation mix and the Southern blot filter (section, 2.3.4.7).

2.3.4.7 Southern transfer of DNA to nylon membranes

2.3.4.7.1 Alkaline transfer method

DNA digested with restriction enzymes (YAC, BAC, PAC, cosmid and plasmid) and PCR products were separated on agarose gels and transferred to Hybond N+™ (Amersham) nylon membrane using the alkaline transfer method (Reed and Mann, 1985). The gel to be blotted was visualised under UV (section, 2.3.4.4.1), photographed, and the DNA size markers were stabbed into the gel with black ink using a needle. The blotting tray contained 0.4 M NaOH transfer solution. The nylon membrane was cut to the size of the gel and was first soaked in distilled water for 5 minutes followed by 0.4 M NaOH solution for another 5 minutes prior to blotting. The gel was placed faced down on the blotting tray followed by the nylon membrane (labelled previously so that the side of membrane with transferred DNA can be orientated), Whatman paper (previously cut to the same size as gel) and paper towels. The DNA in agarose gel was transferred to the prepared filter by capillary action (using the transfer solution) for 16 hours. Next day, the position of wells were marked on the membrane, which was then neutralised in solution containing 0.2 M Tris and 2X SSC for 2 minutes. Membrane was dried at room temperature for ten minutes and was further heated in microwave for 5 minutes at the power level five. Prepared membrane was then used for hybridisation experiments (section, 2.3.4.8). For the DNA (YAC, BAC and PAC) separated by using PFGE, gels were UV nicked using Bio-Rad Gene Linker ™ UV chamber (CA, USA), before the transfer of DNA.

2.3.4.7.2 10X SSC DNA transfer method

The *Pst*I and *Hin*dIII digested genomic and somatic cell hybrid DNA was transferred using 10X SSC transfer solution (Southern, 1975). In this method the gel was denatured and neutralised prior to DNA transfer. The denaturing solution contained 250 ml of 5 M NaCl and 50 ml of 5 M NaOH in a final volume of 500

ml. The neutralisation solution consisted of 150 ml of 5M NaCl and 250 ml of 1 M Tris-HCl, pH 7.5 in a final volume of 500 ml. The agarose gel was first soaked in 500 ml of denaturing solution for 60 minutes, followed by 500 ml of gel neutralising solution for a further 60 minutes. The procedure for transferring the DNA from the gel to the nylon membrane was similar to the alkaline transfer method except that membrane was soaked in 10X SSC (instead of 0.4 M NaOH) for 15 minutes. After transfer the membrane was treated in denaturing solution (300 ml of 0.5 M NaOH) for 30 seconds. The membrane was then neutralised in 1 M Tris-HCl and 20X SSC for 2 minutes and dried as in section 2.3.4.7.1.

2.3.4.7.3 Bacterial colony lifting

The procedure to transfer bacterial clones to nylon membranes was based on a modified version to that described by Grunstein and Hogness, 1975. Bacteria containing a recombinant vector (section, 2.3.3.3.2) were plated on LB agar plates containing relevant antibiotics. After incubation at 37°C, bacterial cells were transferred by placing a nylon membrane disc (Amersham) flat onto the plate surface and leaving for 5 minutes. The filter and plates were marked so that they can be re-aligned accurately. The filters were then transferred to Whatman paper soaked in denaturation solution (1.5 ml NaCl and 0.5 M NaOH) for 7 minutes to lyse the host cells and denature DNA. This was followed by placing filters on Whatman paper soaked in neutralisation solution (3M NaCl and 0.5 M Tris) for 5 minutes. The above step was repeated by transferring filters to a fresh Whatman paper soaked in neutralisation solution for another 7 minutes. Filters were then rinsed in 2X SSC and dried as mentioned in alkaline transfer method.

2.3.4.8 Filter hybridization and washing

Southern filters prepared in section 2.3.4.7 were hybridised based on a modification of Brown (1993). The membranes were prehybridised and hybridised in HYBAID bottles. All hybridisation experiments were carried out in HYBAID orbital midi oven. Prior to addition of labelled probe, nylon filters were wetted in 5X SSC solution, then prehybridised at 42°C for 2-3 hours. The prehybridization solution consisted of 50% (v/v) deionised formamide, 5X SSPE, 2%SDS, 10 ml Denhardt's solution, 10% (w/v) dextran sulphate and 100µg/ml salmon sperm DNA. Radiolabelled probe (1-10 ng/per ml of hybridisation solution) was added to the hybridisation bottle containing hybridisation solution and filter. The filters were allowed to hybridise overnight at 42°C. Nylon filters containing genomic and somatic cell hybrid DNA were prehybridised and hybridised at 65°C for more stringent hybridisation of the probe. Following hybridisation, filters were subjected to washes of varying stringency (Sambrook et al., 1989). Filters were immersed twice in a solution of 2X SSPE and 1% SDS for 10 minutes at 42°C, and then washed in 2X SSPE and 1% SDS (prewarmed at 65°C) for another 30 minutes at 42°C. If high background was still evident, washing was continued in 0.1% SSPE; 1% SDS (prewarmed at 65°C) for another thirty minutes to one hour at 65°C. All the washing steps were carried out in 42°C shaker incubator or a 65°C shaking waterbath. All washed filters were exposed to X-ray films (X-OmatK XK-1 Kodak diagnostic) for visualisation of the hybridised radiolabelled DNA probe. The films were exposed for appropriate time either at room temperature or –70°C.

2.3.4.9 Stripping of nylon filters

Radiolabelled probes, which had been hybridised to HYBOND N+ nylon filters,

were removed by using 0.4 M NaOH solution. This solution was then replaced with neutralisation solution [0.2 M Tris-HCl (pH 7.5), 0.1% SDS and 0.1% SSC]. Both steps were carried out at 42°C for 30 minutes in 42°C shaker incubator.

2.3.5 Polymerase chain reaction (PCR)

Modification of Kogan et al., (1987)

The PCR reactions were performed in a Perkin-Elmer Cetus 480 thermal cycler or a Corbett PCR machine (see section, 2.3.3.4). The PCR reactions typically comprised of 5 µl of 2X PCR mix, 1 µl of 0.5-3 mM $MgCl_2$, 150 ng of each forward and reverse primer, 20 mM BME and 1 unit of Taq polymerase with sterile water to 10 µl. The optimum $MgCl_2$ concentration varied for different primer pairs, although in general satisfactory results were obtained with a concentration of 1.5 mM $MgCl_2$. The PCR reactions were then mixed, briefly spun and overlaid with one drop of paraffin oil. The frequently used thermal cycling program was (94°C for 1 minute, 60°C for 1.5 minutes, 72°C for 1.5 minutes) X 10 cycles, then (94°C for 1 minute, 55°C for 1.5 minutes, 72°C 1.5 minutes) X 25 cycles. A final extension at 72°C for 10 minutes was performed. The PCR products were analysed on 1.5% agarose gels (section, 2.3.4.4.1).

2.3.5.1 PCR primers

The primers designed varied from 20-25 bp in length. The primers contained approximately a 1:1 ratio number of purine and pyrimidine residues. To inhibit the formation of primer-dimers, primers to be used as pairs had different base residues at the 3' end. There were no more than four consecutive residues of the same base in any of the primers designed. The primers used in the study were either designed manually or by using the Primer3 program available online at

(http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi).

2.3.6 *Somatic cell hybrid analysis*

DNA markers, STSs or ESTs at 16p13.3 were localised by PCR and Southern analyses of a chromosome 16 somatic cell hybrid panel. The following chromosome 16 somatic cell hybrids at 16p13.3: 23HA, CY196, CY197, CY182, CY177, CY198 and CY168 (Callen et al., 1995) were used in the mapping experiments. A9 mouse cell line used to construct the hybrids and CY18 a somatic cell hybrid containing an entire chromosome 16 as the only human chromosome were included as control in every experiment for the determination of mouse and human chromosome 16 specific bands. Standard PCR was generally performed using 100 ng of the hybrid DNA in a final volume of 10 µl. PCR conditions were as described in section 2.3.5.

2.3.7 *Screening of high density PAC/BAC human library*

2.3.7.1 Screening of PAC/BAC filters

The high-density dot blot PAC or BAC filter sets were generally used for screening of a PAC or BAC DNA clone corresponding to a particular DNA marker using Southern hybridisation techniques (section, 2.3.4). The PAC library filters and Human Release II BAC library filters were obtained from Genome Systems (St. Louis, Missouri, USA). The RPCI-4 PAC and RPCI-11 (Segment 3 and 4) BAC library filter sets were obtained from PAC/BAC resources (Rosewell Park Cancer Institute, New York, USA). Each set of PAC and BAC filters had an average 3X coverage of the human genome. Each probe was first hybridised to the vector filters in order to eliminate those that contain any homology to vector sequences which, if present, will interfere with the screening of these high-density

grid filters. The PAC/BAC filters were pre-hybridised in 6X SSC, 0.5% SDS and 5X Denhardt's solution and 100 µg/ml denatured salmon sperm DNA at 65^0C for 3 hrs. Probes were labelled with α^{32}P as in section 2.3.4.5, purified (section 2.3.2.2.1) and pre-reassociated (section 2.3.4.6) prior to hybridisation to the filters. Overnight hybridisation with α^{32}P labelled probes were performed at 65^0C in 20 ml of solution containing 6X SSC, 0.5% SDS and 100 µg/ml denatured salmon sperm DNA. Following hybridisation, the membranes were washed in 2X SSC and 0.5% SDS for 5 minutes at room temperature; 2X SSC and 0.1% SDS for 15 minutes at room temperature; 0.1X SSC and 0.5% SDS at 42^0C for 30 minutes; 0.1X SSC and 0.5% SDS at 65^0C for 30 minutes to 1 hr, and finally briefly rinsed in 0.1X SSC at room temperature. Generally the first two washes were required to remove the unincorporated radiolabelled probes. Washed filters were subjected to autoradiography for visualisation of hybridised radiolabelled DNA probe. Interpretations of the results were done as per the protocols of Genome Systems and PAC/BAC Resources.

2.3.7.2 To check for genuine positive PAC/BAC clones

As per the results of autoradiographs obtained from the screening of PAC/BAC filters, positive DNA clones were purchased from Genome Systems or PAC/BAC Resources. These clones were then confirmed to be positive for the particular probe by preparing colony blots (refer section, 2.3.4.7.3). The colony blot filters were labelled with the same probes used to screen the grid filters. Labelling of probes, prehybridisation and hybridisation of colony blot filters were done as per the methods described in sections, 2.3.4.5 and 2.3.4.8.

2.3.7.3 Construction of PAC/BAC contig by restriction fragment analysis

Contig assembly was based on methods described earlier by Giles et al., 1997. The PAC/BAC DNA inserts were digested using infrequently cutting restriction enzymes *Not*I and *Eag*I (New England Biolabs). Restriction digestion reaction conditions were similar as described in (2.3.4.1). The restriction enzyme digested PAC or BAC DNA was then separated by PFGE (refer section, 2.3.4.4.3). The DNA clones were then arranged into a possible contig on the basis of the similarity of the restriction fragment sizes. Fragments representing the extreme ends of individual contigs were used to rescreen the high-density gridded membranes.

2.3.7.4 Isolation of PAC/BAC ends

The 3' and 5' ends of the human inserts of the PACs or BACs were isolated or sequenced to allow determination of the orientation of the constructed contig and for additional cycle of PAC/BAC screening to extend the contig by identification of overlapping clones.

2.3.7.4.1 Isolation of PAC/BAC ends fragments using *Not*I, *Sac*I and *Sac*II restriction enzymes

To obtain the end fragments of the PAC/BAC DNA inserts, the DNA was first digested with a restriction enzyme that frequently cuts the insert but not within the vector. For example, *Sac*I for PAC vector and *Sac*II for BAC vector. The information regarding such enzymes can be obtained from the restriction map of PAC or BAC vector provided in the PAC/BAC manuals of Genome Systems. The digested DNA was then religated using T4 DNA ligase. The restriction digestion and ligation methods were as described in sections 2.3.4.1 and 2.3.3.2

respectively. The religated ends of human inserts were then cloned into XL1 Blue cells (section, 2.3.3.3.2). The DNA was isolated and subjected to *Not*I/*Sac*I and *Not*I/*Sac*II digestion in order to obtain the end fragments. *Not*I releases the insert and *Sac*I and *Sac*II digests human genomic DNA to an average size of 3000-6000 bp. This method was limited to the use of enzymes that only cuts the human inserts of PACs or BACs but not within the vector. Therefore, PAC or BAC end probes were usually generated either by direct sequencing of PAC/BAC DNA inserts or by Bubble PCR methods.

2.3.7.4.2 PAC/BAC end sequencing

Direct sequencing of the ends of human insert of PAC/BAC DNA was performed using Big Dye Cycle sequencing kit (Perkin-Elmer) (refer section, 2.3.12.3). The vector SP6 and T7 promoter sequences flanking the genomic insert were used as sequencing priming sites. The following SP6 and T7 primers were used: SP6 (PAC)- 5' GGCCGTCGACATTTAGGTGACAC 3'; T7 (PAC)- 5' CCGCTAATACGACTCACTATAGGG 3'; SP6 (BAC, pBeloBAC 11)- 5' CGCCAAGCTATTTAGGTGACAC 3' and T7 (BAC, pBeloBAC 11)- 5' GTAATACGACTCACTATAGGG 3'. The sequence generated was used to design PCR primers which allowed amplification of end probes from PAC/BAC DNA for another cycle of PAC/BAC screening to extend the contig by identification of overlapping clones

2.3.7.4.3 Bubble PCR method to generate end probes

The Bubble PCR method used was as described in Gecz et al., 1997, which is a modified version of IRE-bubble PCR approach (Munroe et al., 1994). 1µg of PAC/BAC DNA was digested separately with each of the blunt end cutting

restriction enzymes *Alu*I, *Rsa*I, *Sca*I, *Hae*III and *Pvu*II for ligation to bubble linkers. All reactions were incubated at 37°C for 16 hours. Following digestion, the DNA was extracted using the phenol chloroform method (refer section, 2.3.2.1.1). Bubble anchors were annealed in a total volume of 1 ml with bubble top oligos (5' GAAGGAGAGGACGCTGTCTGTCGAAGGTAAACGGACGA-GAGAAGGGAGAG 3') and the *Not*I-A bubble bottom oligo (5' CTCTCCCTTCTGCGGCCGCAGTTCGTCAACATAGCATTTCTGTCCTCTC CTTC 3') at 3 pmol/µl, 2X SSC and 10 µl of 1M Tris-HCl, pH 7.2. The bubble mix was boiled 15 minutes and slowly cooled to room temperature. The prepared mix was stored in –20°C for further use. The restriction digested PAC/BAC DNA was ligated with 10 µl of 3 pmol/µl annealed bubble linkers in the presence of 10 µl of 10X ligation buffer (MBI) and 2 µl of ligase enzyme (MBI) in a final volume of 100 µl. The reactions were incubated overnight at 4°C. Digested bubble anchor-linkered DNA (10-20 ng) was amplified in 50 µl of PCR reaction consisting of 5 µl 10X buffer, 1 µl of 10 mM dNTP, 2 units of Taq polymerase, 50 pmol of *Not*I-A bubble primer (5'GCGGCCGCAGTTCGTCAACATAGCATTTCT 3') and 50 pmol of SP6/T7 PAC/BAC vector specific primers. 35 PCR cycles were used with each consisting of: 94°C for 30 seconds, 60°C for 30 seconds followed by extension of PCR products at 72°C for 2 minutes. 5 µl of the PCR reaction was analysed on 1.5% agarose gel stained with ethidium bromide and DNA was visualised under UV light. Generated PCR products were purified (section, 2.3.2.2) and either sequenced directly using dye terminator fluorescent kit (Perkin-Elmer) and 3 pmol of individual PCR primers (section, 2.3.12.2) or subcloned into pGEM-T Vector (section, 2.3.3) and sequenced from the vector using M13 forward and reverse dye primers (2.3.12.1). The DNA sequence obtained was used to design primers for

generating insert specific end probes for additional extension of the contig.

2.3.8 *RT-PCR analysis*

The instructions supplied with the Superscript enzyme (Gibco-BRL) were followed for setting of the RT RCR reactions. All reactions were carried out on ice at all times. 2-3 μg of total RNA was mixed with 50 pmol of random hexamers (Perkin Elmer) or 100 pmol of oligo (dT) (Gibco-BRL) in a final volume of 11.5 μl with DEPC treated sterile water. The reaction was heated at 65°C for 5 minutes and placed on ice for 1 minute. Following one-minute incubation on ice, the contents of the tube were spun briefly and 4μl of 5X 1st strand buffer. 2 μl of 0.1 M DTT, 1 μl of 10 mM dNTP and 0.5 μl (20 units) of RNAsin were added. This was incubated at 42°C for 2 minutes, followed by addition of 1 μl (200 units) of the superscript enzyme (Gibco-BRL). After the addition of the superscript enzyme, the reaction was incubated further for 30 minutes at 42°C. The reaction was terminated by incubating at 70°C for 10 minutes and samples were kept at –20°C until needed. A negative control reaction was included in each experiment where the reverse transcriptase enzyme was not added. This allowed the determination of genomic contamination present within RNA sample that may be seen following the PCR step.

The PCR amplification of the reverse transcribed RNA was performed using 2 μl of the reverse transcription reaction, 2.5 μl of 10 X PCR buffer, 0.5 μl of 10 mM dNTPs, 0.75 μl of 50 mM $MgCl_2$, and 1 μl of each forward and reverse primer (150 ng/μl) in a final volume of 20 μl. The reactions were overlaid with one drop of paraffin oil and incubated at 96°C for 5 minutes. Following 5 minutes

incubation, 1 µl (0.5 units) of Taq DNA polymerase was added to all tubes, including those tubes without reverse transcriptase. While the Taq DNA polymerase enzyme was added to PCR tubes the temperature of the thermal cycler was maintained at 80°C. After the addition of Taq DNA polymerase enzyme, the reactions were incubated at 94°C for 4 minutes. Then 35 PCR cycles were performed each consisting of 94°C for 1 minute; 60°C for 2 minutes; 72°C for 3 minutes. This was followed by a final extension of the amplified products at 72°C for 7 minutes. 10 µl of the RT-PCR reaction was visualised on 1.5% agarose gel (containing ethidium bromide) in 1 X TBE. The remaining reaction was purified (section, 2.3.2.2) and subcloned into pGEM-T Vector (section, 2.3.3) for sequencing purposes (sections, 2.3.12.1 and 2.3.12.2). To test the success of the cDNA synthesis, esterase D primers (forward primer) 5' GGAGCTTCCCCAACTCATAAATGCC 3' (423-447; GenBank Accession Number M13450) and (reverse primer) 5' GCATGATGTCTGATGTGGTCAGTAA 3' (875-851; GenBank Accession Number M13450) were used as a positive control for each RT-PCR experiment. These primers amplified a 452 bp product from the cDNA template only.

2.3.9 *Multiple Tissue Northern Blot (MTN™) analysis*

Commercial Northern blot of multiple adult human tissue mRNAs was obtained from Clontech, Palo Alto, USA and hybridised according to protocols supplied with the Clontech MTN blot (User manual, PT 1200-1). The filter contained 2 µg of poly A^+ RNA derived from heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreatic sources. The Northern blot was prehybridised and hybridised using ExpressHyb™ hybridisation buffer (Clontech). Prior to addition of ExpressHyb™ hybridisation buffer (prewarmed at 65°C) to the filter, 37.5 µl of

20 mg/ml Salmon sperm DNA (previously boiled to 100°C and cooled on ice for 1 minute) was added. Prehybridisation of the membrane was carried out for 2 to 3 hours at 65°C and hybridisation proceeded overnight at 65°C. Methods of radiolabelling and pre-reassociation of DNA probe was same as described in sections 2.3.4.5 and 2.3.4.6 respectively. Following hybridisation, the filter was rinsed at room temperature for 30-40 minutes with continuous agitation in wash solution (2X SSC and 0.05% SDS), which was replaced several times. Counts were monitored after each rinse. If the counts were still high (greater than 50 cpm), the membrane was then transferred to a solution (0.1X SSC and 0.1% SDS, prewarmed at 50°C) for 40 minutes with continuous shaking at 42°C with change of fresh solution after every 10 minutes. Counts were again checked after every 5 minutes and the washing of the membrane was continued until counts were at a level of less than 10 cpm. Filters were sealed in a plastic bag and exposed to X-ray film at –70°C with two intensifying screens. Subsequently the probe was removed from the Northern blot by incubating the blot in sterile water containing 0.5% SDS (previously boiled to 100°C for 10 minutes) for 30 minutes at room temperature. The stripped membrane was stored at –20°C until needed.

2.3.10 Human RNA Master Blot™ analysis

Human RNA Master blot was obtained from Clontech, Palo Alto, USA and hybridised according to protocols supplied with the Clontech RNA master blot (User manual, PT 3004-1, catalogue number 7770-1). The Human RNA blot is a positively charged nylon membrane and allows the determination of the relative level expression of a target mRNA in different tissues and developmental stages. It contains poly A^+ RNAs from fifty different human tissues with (100-500 ng/ 1-mm dot) poly A^+RNA normalised with respect to several housekeeping genes.

The method of dot blot hybridisation was similar to MTN blot hybridisation except that here 15 ml of ExpressHyb mix was used. Following hybridisation, the filter was rinsed four times in 2X SSC and 1% SDS (prewarmed at 65°C) for 20 minutes at 65°C. If high background was still evident, the membrane was further washed twice in second wash solution (0.1X SSC and 0.5% SDS, prewarmed at 55°C) for 20 minutes at 55°C or until the counts were at a level of less than 10 cpm. Autoradiography and stripping of dot blot filter was performed as (2.3.9).

2.3.11 *Fluorescence in situ hybridization(FISH)*

The FISH analysis involved in the study was initially undertaken by Elizabeth Baker and later was mainly carried out by Helen Eyre (Department of Cytogenetics, WCH). The method used is briefly described. The DNA clones (plasmids, cosmids, PACs, BACs, or YACs) were labelled by nick translation with biotin-14-dATP (Gibco-BRL, Life Technologies, Gaithersburg, USA). FISH was performed as previously described (Callen et al., 1990a), except that the chromosomes were mounted prior to analysis in antifade containing propidium iodide (as counterstain) and DAPI (for chromosome identification). Images of metaphase chromosome preparations were captured using the Chromoscan Image Collection and Enhancement System (Applied Imaging Corporation, New Castle, UK). FISH signals and DAPI banding patterns were merged for figure preparation.

2.3.12 *DNA sequencing*

DNA sequencing involved in the present study was carried out using Dye Primer, Dye Terminator and BigDye™ Terminator Cycle Sequencing Ready Reaction kits (ABI-Perkin Elmer, Warrington, UK) according to the manufacturer's protocols.

In all the cases (Dye Primer, Dye Terminator and Bigdye Fluorescent Chemistry) the sequencing reactions were cleaned and sent for subsequent gel run on ABI377 sequencer to IMVS (Institute for Medical and Veterinary Science), Adelaide, Australia. The sequence obtained was analysed using the Laser Gene DNA software package (DNAstar, Madison, USA).

2.3.12.1 Dye Primer Cycle Sequencing

cDNA clones and PCR products cloned in pGEM-T vector were sequenced using ABI Prism™ Dye Primer Fluorescent Chemistry. Four separate tubes were needed for each DNA template. The reaction volumes of reagents are given in Table 2.2. After setting the Dye Primer sequencing reaction, the reaction mixture was overlaid with one drop of paraffin oil.

Table 2.2 Working protocol for Dye Primer sequencing reaction.

Reagent	A (μL)	C (μL)	G (μL)	T (μL)
DNA Template (200 ng/μl)	1	1	2	2
Ready Reaction Premix	4	4	8	8

The PCR thermal cycling program used for Dye Primer Cycle Sequencing was: 15 PCR cycles 95°C for 30 seconds, 55°C for 30 seconds, 70°C for 1 minute followed by another 15 PCR cycles of 95°C for 30 seconds and 70°C for 1 minute.

The sequencing reaction was then purified as per the manufacturer's protocol. The PCR reaction for each of the four tubes (A, C, G, and T) were combined and transferred to eppendorf tube containing 100 μl of 100% ethanol and 1.5 μl of 3 M sodium acetate. The reactions were mixed thoroughly by brief vortexing. The

tubes were placed on ice for 30 minutes and then spun in a microcentrifuge for 30 minutes at maximum speed. The supernatant was discarded and the pellets were washed with 70% of ethanol to remove the salts from the pellet. DNA pellets were dried at room temperature for 15 to 30 minutes.

2.3.12.2 Dye Terminator Cycle Sequencing

Plasmid inserts and PCR products (previously cleaned using methods described in section 2.3.2) were sequenced using specific sequencing primers and Dye Terminator Cycle Sequencing Kit. The reaction volumes of reagents are given in Table 2.3. The PCR cycle conditions used for dye terminator cycle sequencing reactions were as follows: 25 PCR cycles of 96°C for 30 seconds, 50°C for 15 seconds, and 60°C for 4 minutes.

Table 2.3 Working protocol for Dye Terminator sequencing reaction.

Reagent	Quantity
Template	
Double stranded DNA (200 ng/µl)	2.5 µl
PCR product (30 ng/µl)	6 µl
Primer	3.2 pmol (2 µl of 10 ng/µl)
Terminator Ready Reaction Mix	8 µl (4 µl of dye terminator reaction mix + 4 µl of *half* Term™ mix)

Final reaction volume was made to 20 µl with sterile water and overlaid with a drop of paraffin oil.

The dye terminator sequencing reactions were cleaned as per the manufacturer's instructions in order to remove the excess dye terminators using ethanol/sodium acetate precipitation method. The entire contents of each extension were

transferred to a 1.5 ml microcentrifuge tube containing 2.0 µl of 3 M sodium acetate (pH, 4.6) and 50 µl of 95% ethanol. The tubes were vortexed and placed on ice for 15 minutes to precipitate the extension products. Tubes were then spun in a microcentrifuge for 30 minutes at maximum speed. The supernatant was discarded, pellets were rinsed with 250 µl of 70% ethanol and dried for 30 minutes to one hour at room temperature.

2.3.12.3 BigDye™ Terminator Cycle Sequencing Kit

The PAC/BAC DNA was sequenced using Bigdye Terminator Cycle Sequencing Kit and SP6 and T7 sequencing primers (section, 2.3.7.4.2) to generate the PAC/BAC end sequences. The reaction volumes of reagents are given in Table 2.4. The Bigdye sequencing reactions were denatured at 95°C for 5 minutes prior to PCR run followed by 80 PCR cycles of 95°C for 30 seconds, 50°C for 20 seconds and 60°C for 4 minutes. The Bigdye sequencing reactions were cleaned as in section 2.3.12.2.

Table 2.4 Working protocol for Bigdye sequencing reaction.

Reagent	Quantity
PAC/BAC DNA Template (1.5 µg/µl)	1 µl
Sequencing Primers (SP7 and T7, 50pmol/µl)	1 µl
BigDye™ Terminator Mix	8 µl (4 µl terminator mix + 4 µl *half* BD™ mix)

The final reaction volume was made 20 µl with sterile water and overlaid with a drop of paraffin oil.

Physical Mapping At 16p13.3 Identifies A Novel Proline Rich Gene, C16*orf*5

Chapter 3

3.1 Introduction

Balanced chromosomal translocations occur in humans with a frequency of approximately 1 in 800 individuals and are usually associated with a normal clinical phenotype (Warburton, 1991). However, the risk of serious congenital anomaly is estimated to be 6.1% (n=163) for *de novo* reciprocal translocations (Warburton, 1991). A somatic cell hybrid panel of chromosome 16 was constructed from various translocations and interstitial deletions involving this chromosome, which were ascertained in different cytogenetic laboratories (Callen, 1986). Included in this panel were two *de novo* translocations t(1;16) and t(14;16) associated with mental retardation and epilepsy, which were characterised in the present study. The mouse/human somatic cell hybrids isolated from the cell lines of these two patients containing the der(16)t(1;16) and the der(16)t(14;16) were designated as CY196 (Callen et al., 1990b) and CY182 (Callen et al., 1995) respectively. The hybrid breakpoints included in the chromosome 16 somatic cell hybrid panel were mapped in detail (Callen et al., 1992, 1995). These data revealed that the hybrid breakpoints CY196 and CY182 reside in close vicinity on chromosome 16 at 16p13.3. It was proposed that the mental retardation and epilepsy in the two patients with the *de novo* balanced translocations was related to the disruption of a gene at or near the 16p13.3 breakpoints of the t(1:16) and t(14;16).

To identify the disrupted transcript at 16p13.3, it was first necessary to generate a more detailed physical map of the breakpoint region at 16p13.3 by mapping STSs and ESTs in the vicinity of the breakpoints of the two translocations. The hybrids CY196 and CY182, which contained the der(16) of the two translocations were used for this physical mapping. This would allow the identification of a DNA

probe in the vicinity of the breakpoints, which will be used to screen PAC/BAC libraries to allow the construction of a physical map. FISH analysis will be done with these DNA clones on the metaphase chromosomes of the two patients in order to identify the DNA clone spanning the 16p13.3 breakpoints. This preliminary physical mapping in the vicinity of the chromosome breakpoints in CY196 and CY182 identified an EST WI-16589, which was mapped on the GB4 radiation hybrid panel. A BLAST search of dbEST database with the sequence of WI-16589 identified two overlapping cDNA clones AA070715 and H07909 that together represent a novel proline rich gene, which was notated C16*orf*5. This gene is highly expressed in the brain. Because of this expression pattern and the localization immediately distal to CY196-CY182 hybrid interval it was a possible candidate for a gene, which would be disrupted by the two translocations at 16p13.3 involved in mental retardation and epilepsy.

This Chapter presents the preliminary physical mapping data in the vicinity of the 16p13.3 breakpoints contained in CY196 and CY182, and identification, localization, characterization and expression studies involving the C16*orf*5 gene.

3.2 Methods

Only the methods used specifically in this Chapter are discussed in detail. The general molecular genetic techniques of DNA analysis are described in Chapter 2.

3.2.1 Somatic cell hybrid analysis

Localization of the DNA probes on the short arm of chromosome 16 was as described in section 2.3.6 by PCR analysis of somatic cell hybrids with breakpoints at 16p13.3 (Callen et al., 1995). Included in this panel were the

hybrids CY196 and CY182, which were derived from the *de novo* balanced translocations ascertained in the two patients with mental retardation and epilepsy. The A9 mouse cell line used to construct the hybrids, and CY18 a somatic cell hybrid containing an entire chromosome 16 as the only human chromosome, were included as controls in every experiment for the determination of mouse and human chromosome 16 specific bands. The primer sequences of the DNA probes localized in the vicinity of the hybrid interval defined by CY196 and CY182 are presented in Table 3.1. DNA from the hybrids at 16p13.3 was analyzed by PCR and Southern blot (2.3.4) analyses to localize the C16*orf*5 transcript.

3.2.2 Generation of C16orf5 cDNA sequence

The end sequences for both of the cDNA clones AA070715 and H07909, which represents the C16*orf*5 transcript, were available in the GenBank. This enabled the design of primers (section 2.3.5.1) to generate the complete cDNA sequence. The cDNA sequence was generated using Dye Terminator fluorescent chemistry and gene specific primers as described in section 2.3.12.2. Sequencing samples were analyzed on model ABI377 automated sequencer (Applied Biosystems).

3.2.3 5' RACE (rapid amplification of cDNA end) analysis

The 5' end of the C16*orf*5 transcript was confirmed by 5' RACE on polyA$^+$ mRNA from fetal brain tissue, using a 5' RACE Kit (Gibco BRL, Life Technologies, Gaithersburg, USA). All the components were supplied with the kit except the gene specific primers (GSP). The sequence of the GSP primers used in 5' RACE is presented in the section 3.3.4. 500 ng of polyA$^+$ fetal brain mRNA was reverse transcribed with 2.5 pmoles (approximately 10 to 25 ng) of GSP1 at 42°C for 50 minutes, followed by a incubation at 70°C for 15 minutes to terminate

the reaction. After the addition of 1 µl of the RNAase mix, the tubes were left for a further 30 minutes at 37°C. The first strand cDNA was purified using kit supplied Glass Max DNA isolation spin cartridge. The cDNA was eluted from the column with 50 µl of sterile water (preheated to 65°C) by centrifugation at 13,000 rpm for 20 seconds. The purification of cDNA was followed by the addition of homopolymeric tail at the 3'-end of the cDNA using TdT and dCTP. For this 10 µl of the purified sample was added to 6.5 µl of DEPC treated water, 5 µl of 5X tailing buffer, and 2.5 µl of 2mM dCTP and the reaction contents were incubated for 3 minutes at 94°C followed by a one minute incubation on ice. After the addition of 1 µl of the TdT mix, the tubes were incubated at 37°C for 10 minutes. The TdT was then heat inactivated for 10 minutes at 65°C. The dC-tailed cDNA was subjected to PCR using a GSP2 primer and the abridged anchor primer (AAP, which was supplied with 5'RACE Kit). The 3' end of the anchor primer contained a sequence complementary to the homopolymeric tail. This step was followed by a nested PCR using a GSP 3 primer and abridged universal amplification primer (AUAP, was also supplied with the 5'RACE Kit). The AUAP contains a restriction endonuclease site sequence at the 5' adapter region homologous to the adapter region of the anchor primer.

The sequence of the AAP and AUAP primers are as follows: AAP 5'- GGC CAC GCG TCG ACT AGT ACG GGI IGG GII GGG IIG-3'; AUAP 5'- GGC CAC GCG TCG ACT AGT AC-3'. The anchor primer was designed with selective placement of deoxyinosine residues in the poly (dG) portion. This selective placement of deoxyinosine residues in the 3'region of the anchor primer maintains low stability on the 3' end of the primer, creates a melting temperature (Tm) for the 16 basepair anchor region (66°C) which is comparable to that of a typical 20-

mer primer with 50% GC content, minimizes priming at internal C-rich regions of cDNA and maximizes specific priming for the oligo-dC tail (Instruction Manual of the 5' RACE system, Version 2.0, catalogue number 18374-058, Gibco BRL, Life Technologies).

The PCR reactions were performed as per described in the manufacturer's instruction manual. However, before the addition of the Taq DNA polymerase, the tubes containing all components, were overlaid with a drop of paraffin oil and incubated at 94°C for one minute. Then 0.5 µl of Taq DNA polymerase (5 units/µl) was added followed by 35 PCR cycles of 94°C for one minute, 60°C for two minutes, 72°C for three minutes, with a final incubation of 72°C for seven minutes. Following PCR amplification of 5' RACE products, 15 µl of the PCR products were analyzed on 2.5% (w/v) agarose gel, and the remaining sample was cloned into the pGEM-T vector (section 2.3.3), subsequently the cloned 5' RACE products were sequenced using Dye Primer fluorescent chemistry (section 2.3.12.1).

3.2.4 Partial genomic characterization of C16orf5

To determine the partial genomic structure of the C16orf5 gene, DNA was prepared (2.3.1.3) from the PAC 102J11 shown to contain the gene. The DNA was digested separately with the blunt end cutting restriction enzymes *Alu*I, *Rsa*I, *Sca*I, *Hae*III and *Pvu*II and tentative exonic boundaries were generated as described in section 2.3.7.4.3. The resulting PCR products were purified (2.3.2.2.1) and then directly sequenced using Dye Terminator cycle sequencing kit (2.3.12.2) and the gene specific primers.

3.2.5 Database homology searches

The BLAST search tool (Altschul et al., 1990, 1994) at NCBI was used for the C16*orf*5 cDNA nucleotide sequence homology searches with sequences deposited in GenBank non-redundant and EST databases. Further analysis of C16*orf*5 cDNA nucleotide sequence involved PROSITE database (Hofmann et al., 1999) and PSORTII (Nakai and Horton 1999).

3.2.6 Expression studies involving C16orf5 gene

Size and expression of the C16*orf*5 transcript was evaluated using a Multiple Tissue Northern (MTN^{TM}) Blot and RNA Master Blot, both obtained from Clontech, as per the methods described in section 2.3.9 and 2.3.10 respectively. The blots were probed with α^{32}P-dCTP labelled insert of cDNA clone AA070715. The cDNA was digested with restriction enzymes to isolate the insert (2.3.4.2). The insert of cDNA was then purified (2.3.2.2.2) and labelled as described in section 2.3.4.5 followed by the pre-reassociation of the probe as per the method described in section 2.3.4.6.

3.2.7 PAC/BAC DNA clones

The PAC and BAC DNA clones spanning the C16*orf*5 transcript were identified by the hybridisation screening of the two different human PAC libraries (Genome Systems, St Louis, Missouri, and PAC/BAC Resources, Buffalo, New York) and a human BAC library RPCII-11 Segment 3 from PAC/BAC Resources. Each set of PAC and BAC filters had an approximate three times coverage of the human genome. These libraries were screened with the α^{32}P-dCTP labelled inserts of cDNA clones AA070715 and H07909 using the protocol described in section

2.3.7.1. The positive DNA clones were purchased from the Genome Systems and PAC/BAC Resources. These clones were confirmed as real positives for the cDNA clones AA070715 and H07909 by preparing colony blots (section 2.3.4.7.3). The real positive clones were grown in LB medium (2.2.6.1) containing 50 μg/ml of kanamycin (PACs) and 30 μg/ml of chloramphenicol (BACs). The cultures were grown overnight shaking at 37°C. DNA was isolated using Qiagen tip-100 columns as described in section 2.3.1.3. The clones were then sized and restriction mapped (2.3.7.3) to arrange the DNA clones in a contig.

3.2.8 Fluorescence in situ *hybridization (FISH)*

FISH analysis of cDNA clones AA070715 and H07909 and PAC 102J11 was performed as described in section 2.3.11 by Elizabeth Baker and Helen Eyre (Department of Cytogenetics and Molecular Genetics, WCH, Adelaide). The DNA clones were labelled by nick-translation with biotin-14-ATP and hybridised *in situ* to the metaphase chromosomes.

3.3 Results

3.3.1 Physical map in the vicinity of the hybrid breakpoints CY196 and CY182 at 16p13.3

At the beginning of the project there were no DNA markers localized within the hybrid interval defined by CY182 and CY196 (Callen et al., 1995) at 16p13.3, therefore, this region was mapped in detail to identify a DNA probe in the vicinity of the translocation breakpoints at 16p13.3.

Table 3.1 Primer sequences for STSs/ESTs physically mapped on the short arm of chromosome 16 at 16p13.3.

DNA markers	Primer sequences (5'-3')		PCR product size (base pairs)
	Forward primer	*Reverse primer*	
WI-18474	ctggactggggctatgaaac	gtggcactgagctatgagtcc	105
WI-14401	ttattgtcttctccttcgacagc	tctttcaccacgtgtgaactg	131
WI-13954	aaagtttacgaactcgtgttgg	aaaaccctggaggtttccc	150
WI-11319	gcagaaaagaaatttattaccaagc	gatgaggaactcaaaatccagg	129
SGC34290	ccgatctagttttagcacagagc	ctgtgcttcccctcctgg	130
WI-12224	tttacgtaccattcaattcaccc	ggggaacatggaatctcctt	134
D16S2965	ttatgacttcaagaatgtcatcctg	gtgtcttggtattttcacatctcg	277
D16S3327	agtctttggccccacactc	tgtcttcaggttaaattggaagg	143
DNL1	attgttcccctgcatgcggc	tgctcacttcagcatcacct	900
WI-21938	gtggcaggatacagtggtcc	caaagggtcccccttgtg	264
IB1814	gatcctgaaacacatggaaa	gttcccccacttacactgtt	207
WI-16589	ttcagtccagactctttcccc	cctgactccagacaacttaccc	127
KB#1	gtgtccaccaaggcagttct	gctgacaggctgttgctgt	100
D16S423	aaacaggcttgaaagtctctg	ttgtcttctgtccacttacaca	292
D16S3232	aatgcaatatacaaaaactcaccc	attttaaacaattttgtcatcacca	101
AFM339xg1	cttcactgggggtaggg	agtaaaaacgtccttgcatc	210
AFM288zf9	ttgtgcagacttgtccat	cagagccagagtgagagg	198
SGC32535	gtttcccaggctgatctcaa	aggttgaaaagtaaaaattgggc	133
GGAA24H01	ggacctctctaggcaagacc	tccactcattgaacaatactgc	214
c72D12	ctctgtgatgacaggaactgtg	atctctggatggagacaccatc	167
A002D07	tagtattccatggtgtatatgt	ctataaagacacatgcaca	132

The STSs and ESTs present in the vicinity of the breakpoint were obtained by accessing the radiation hybrid map of the Whitehead Institute and the CEPH (Centre d'Etude du Polymorphisme Humain) genetic map of chromosome 16. A list of the primer pairs for STSs and ESTs from this in silico analysis, which were likely to map near the chromosome breakpoints of CY196 and CY182 is presented in Table 3.1. The DNA markers were mapped on chromosome 16 by PCR analysis of the somatic cell hybrids 23HA, CY182, CY177, CY196, CY197, CY198, CY168, which all had breakpoints at 16p13.3. The results of physical mapping of DNA markers on short arm of chromosome 16 at 16p13.3 are depicted in Table 3.2 and are summarised in Figure 3.1.

The order of the hybrids presented in Table 3.2 is as described in the earlier reported map of chromosome 16 (Callen et al., 1995). The data shown in Table 3.2 demonstrate that the markers WI-18474 to D16S423 are located distal and the markers from D16S3232 to A002D07 are located proximal to the hybrid interval defined by CY182 and CY196. The DNA markers D16S3232, AFM339xg1 and AFM288zf9 were identified to map within the hybrid interval defined by CY196 and CY197, thereby separating the chromosome breakpoints in these two hybrids. There were no markers identified earlier separating these two breakpoints in the previously reported map of chromosome 16 (Callen et al., 1995). The results presented in the Table 3.2 refined the order of the hybrids breakpoints CY177, CY182-CY196, CY197, from that reported earlier in Callen et al., (1995) to pter-CY196-CY197, CY177, CY182 (Figure 3.1).

Table 3.2 Physical mapping of probes to the short arm of chromosome 16 at 16p13.3 by PCR analysis of mouse/human somatic cell hybrids

DNA Markers	Somatic cell hybrids at chromosome 16						
	CY18	*23HA*	*CY177 CY182*	*CY196*	*CY197*	*CY198*	*CY168*
WI-18474	+	+	-	-	-	-	-
WI-14401	+	+	-	-	-	-	-
WI-13954	+	+	-	-	-	-	-
WI-11319	+	+	-	-	-	-	-
SGC34290	+	+	-	-	-	-	-
WI-12224	+	+	-	-	-	-	-
D16S2965	+	+	-	-	-	-	-
D16S3327	+	+	-	-	-	-	-
DNL1	+	+	-	-	-	-	-
WI-21938	+	+	-	-	-	-	-
IB1814	+	+	-	-	-	-	-
WI-16589	+	+	-	-	-	-	-
KB#1	+	+	-	-	-	-	-
D16S423	+	+	-	-	-	-	-
D16S3232	+	+	-	+	-	-	-
AFM339xg1	+	+	-	+	-	-	-
AFM288zf9	+	+	-	+	-	-	-
SGC32535	+	+	+	+	+	+	+
GGAA24H01	+	+	+	+	+	+	+
c72D12	+	+	+	+	+	+	+
A002D07	+	+	+	+	+	+	+

+ indicates the probe present in the hybrid line; -, probe is absent in hybrid line. The order of the hybrids presented is as described in Callen et al., 1995. The EST shown in red identified the C16*orf*5 gene. The markers shown in green separated the chromosome breakpoints contained in CY196 and CY197. These markers were absent in the hybrid cell line CY182, which refined the order of hybrid breakpoints CY177, CY182-CY196, CY197 (reported earlier in Callen et al., 1995) to pter-CY196-CY197, CY177, CY182-CY198, CY168, where comma indicates the unknown order and a dash known order. The results from this Table are summarized in Figure 3.1.

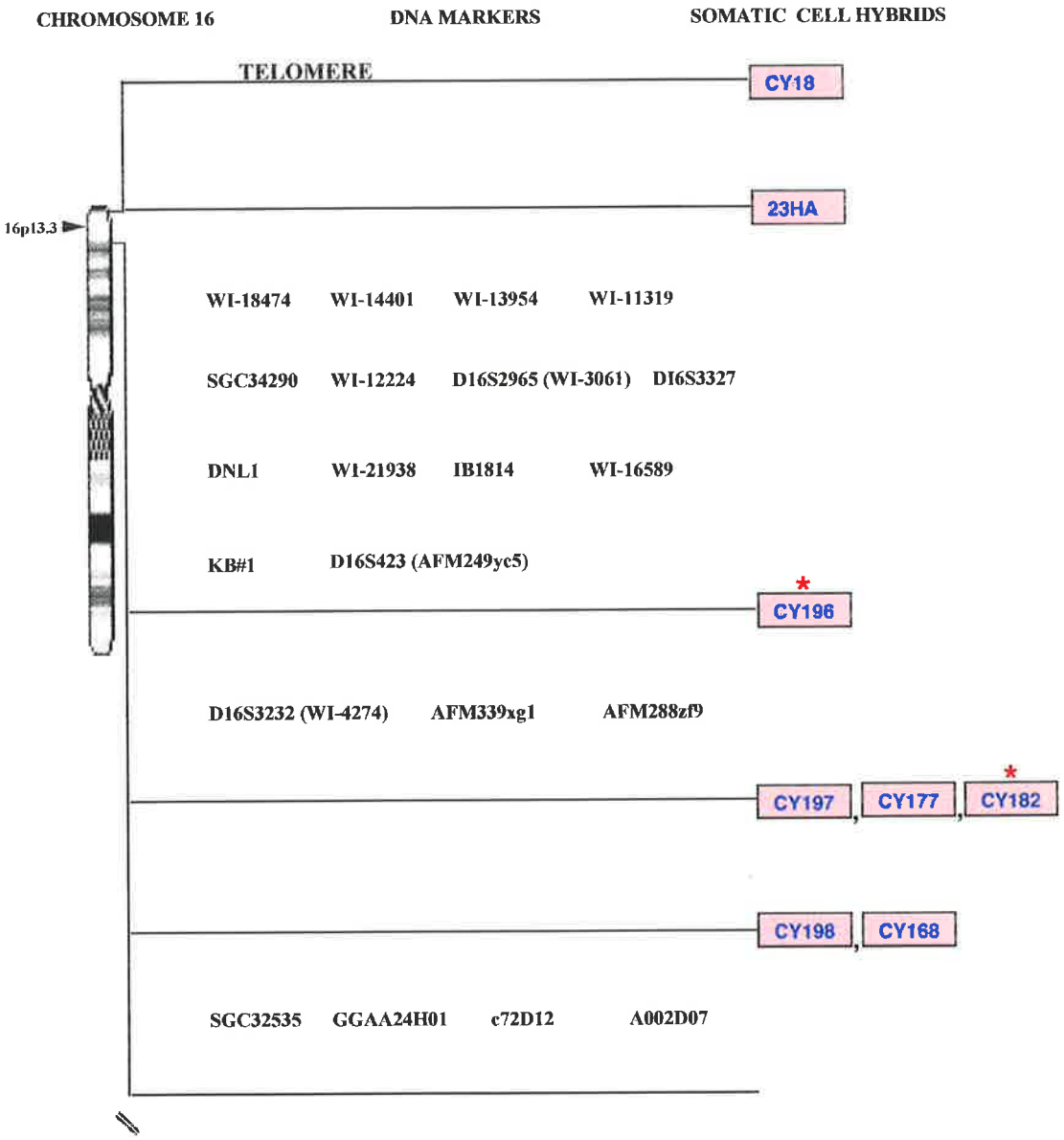


Figure 3.1 Physical mapping of probes to the short arm of chromosome 16 in the vicinity of the 16p13.3 breakpoints contained in CY196 and CY182. The somatic cell hybrids 23HA, CY177, CY182, CY196, CY197, CY198 and CY168 mapped at 16p13.3 (Callen et al., 1995) were used for the mapping experiments. The CY18 somatic cell hybrid that contained the entire chromosome 16, was used to identify the chromosome 16 specific bands. The CY196 and CY182 hybrid breakpoints characterised in the present study are depicted by asterisks. The orientation of the chromosome breakpoints in the hybrids CY177, CY182-CY196, CY197 as reported in Callen et al., 1995 was refined to CY196-CY197, CY177, CY182, where comma indicates order unknown and a dash known order. These data separated the chromosome breakpoints of CY196 and CY197. This figure summarises the results from Table 3.2. Some of the markers shown within brackets had the D numbers available that are indicated in preference.

It was not possible to physically map the DNA markers KIAA0421 (forward primer 5' acggtgcttggtggagtatc 3; reverse primer 5' cagtccggagtaacactggc 3') and WI-22537 (forward primer 5' caccatcacatttcaagtttcc 3'; reverse primer 5' ccttttaatgtgggcagtgg 3') because they amplified a similar size mouse specific band from the mouse line A9. And the marker SGC34289 (forward primer 5' aagggaaaggtcttgaataagtaca 3'; reverse primer 5'tctaaggggttcaagtagtaatg 3') did not amplify a band from the somatic cell hybrid CY18, which suggested that it is not present on chromosome 16.

3.3.2 Identification of C16orf5 gene

An EST WI-16589 was mapped on the GB4 radiation hybrid panel at 16p13.3 (http://www.ncbi.nlm.nih.gov/genemap98) and also has been located in the adjacent distal somatic hybrid interval to hybrid breakpoints CY182 and CY196 at 16p13.3 (Table 3.2). A blast search of the dbEST database with the sequence of WI-16589 identified two overlapping cDNA clones AA070715 and H07909, which together represents the novel proline rich gene C16*orf*5.

3.3.3 Localization of C16orf5 gene

The EST WI-16589 (Table 3.2), which was present at the 5' end of the cDNA clone H07909, was localised in the adjacent hybrid interval to hybrid breakpoints CY196 and CY182 by PCR analysis of the somatic cell hybrids at 16p13.3 (Figure 3.2 A). To also PCR map the 3' end of the gene with respect to the 16p13.3 breakpoints contained in CY196 and CY182, primers were designed from the 3' end sequence of cDNA clone AA070715 (which was available in dbEST database) and were designated as KB#1 forward and reverse (Table 3.1).

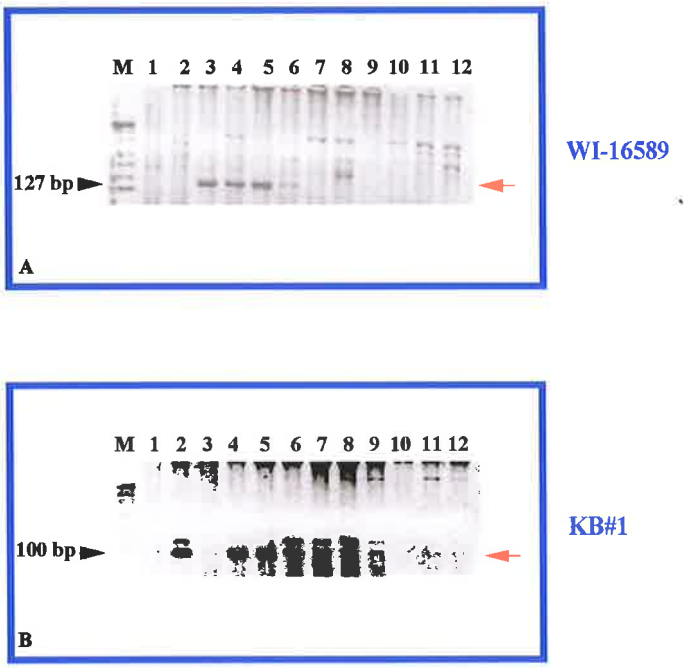


Figure 3.2 Localization of C16*orf*5 by PCR analysis of somatic cell hybrids at 16p13.3. **(A)** PCR mapping of of WI-16589 by somatic cell hybrid analysis. Samples: M1) pUC19 DNA marker, 1) no DNA control, 2) A9 mouse DNA, 3) human blood bank DNA, 4) CY18, 5) 23HA, 6) CY194, 7) CY177, 8) CY182, 9) CY196, 10) CY197, 11) CY198, 12) CY168. **(B)** PCR mapping of KB#1 by somatic cell hybrid analysis. The order of samples loaded is same as in gel A except that in this gel lanes 1-3 represent no DNA control, total human DNA and A9 mouse DNA respectively. The red arrows depicts the chromosme 16 specific bands. The sizes of PCR products are indicated.

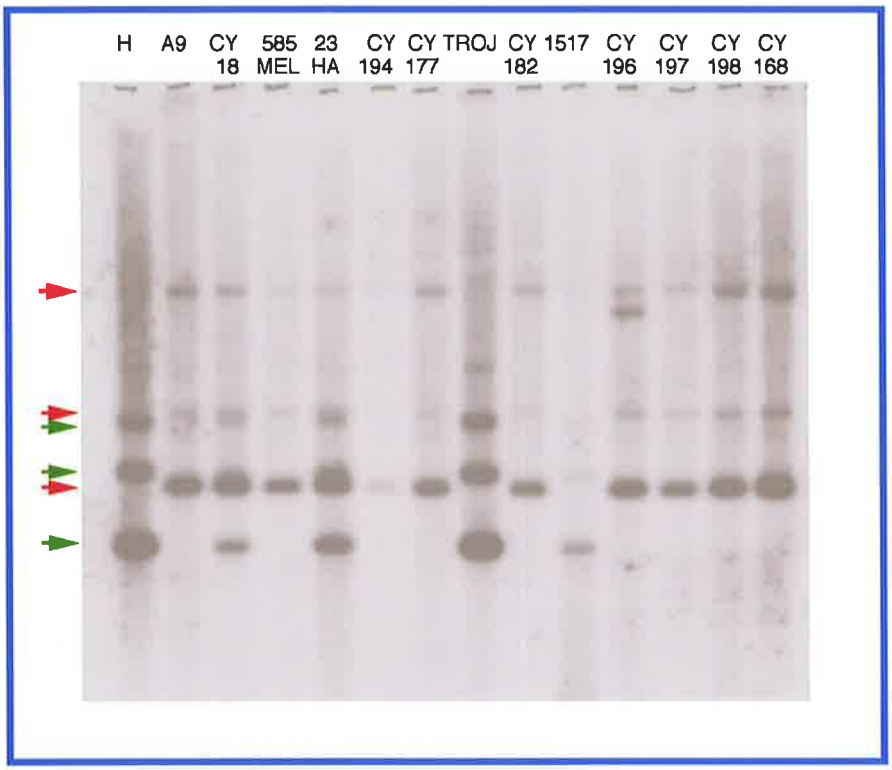


Figure 3.3 The somatic hybrids CY194, CY177, CY182, CY196, CY197, CY198 and CY168 at 16p13.3 were used for the localization of the C16*orf*5 transcript. TROJ and 1517 fibroblast cell lines contained chromosome 16 translocations represented in the hybrids CY177 and CY182. The 1517 cell line was used as as control to detect the genomic fragment spannig the CY182 hybrid breakpoint. A9 and 585-MEL cell lines were included as controls to eliminate rodent specific bands in the somatic cell hybrids. Human genomic DNA in H lane was used to determine the human specific bands. The somatic cell hybrid CY18 was used to determine the chromosome 16 specific bands. The red arrows in the figure depict the mouse specific bands and the green arrows depicts the human specific bands. The *Pst* I digested genomic DNA was probed with the inserts of cDNA clones AA070715 and H07909. The isolation and purification of inserts of cDNA clones was done by the candidate and subsequent preperation and probing of Southern membrane containing *Pst* I digested genomic DNA was carried out by Scott A. Whitmore, Department of Cytogenetics, WCH, Adelaide. The data indicates that C16orf5 gene localizes in the adjacent distal hybrid interval defined by CY182 and CY196.

However, from the PCR mapping result (Figure 3.2 B) it was difficult to interpret whether the 3' end of this gene lies distal or proximal to the 16p13.3 translocation breakpoints because along with the chromosome 16 specific band amplified in the hybrids CY194, CY177and CY182 an extra band was also amplified in these hybrids and the chromosome 16 specific band observed in CY196 hybrid cell line was faint. Therefore, the localisation of C16orf5 gene was also confirmed by Southern analysis of a human/rodent somatic cell hybrid panel (Callen et al., 1995) using cDNA clones AA070715 and H07909. *Pst*I and *Hin*dIII digested somatic cell hybrid DNA was blotted and hybridised with the inserts of cDNA clones (Figure 3.3). The Southern analysis also localised the gene in the adjacent hybrid interval to hybrid breakpoints CY196 and CY182 at 16p13.3.

3.3.4 cDNA sequence of C16orf5

The end sequences for both of the cDNA clones AA070715 and H07909 were available in dbEST database, which enabled the design of primers for sequencing the cDNA inserts. The insert lengths for overlapping cDNA clones AA070715 and H07909 as mentioned in the database were 2253 bp and 1719 bp respectively. The cDNA inserts were isolated using restriction enzymes *Eco*RI and *Xho*I for cDNA AA070715 and *Not*I and *Hin*dIII for cDNA H07909. The cDNA sequence generated with the help of these overlapping cDNA clones is presented in Figure 3.4. The 5'RACE analysis was performed on poly A^+ mRNA from fetal brain tissues to confirm the 5' end of the gene. The following nested gene-specific primers were used to generate 5' RACE products: GSP#1 (nt 362-346) 5'CTT TTC TCT TCC AGA AG 3', GSP#2 (nt 309-290) 5'CTC GCT GGA CAT CTT CGC TG 3' and GSP#3 (nt 296-277) 5' TTC GCT GCT TCT CCT CAA AT 3' (Figure 3.4). Following amplification of a 5'RACE product (Figure 3.5), the

product was cloned and sequenced, which generated a further 193 bp sequence at the 5' end of the gene (Figure 3.4). The complete cDNA sequence presented in Figure 3.4 represented a novel gene, which was designated as C16*orf*5 (chromosome 16 open reading frame #5) and was assigned a GenBank accession number, AF131218.

3.3.5 Characterization of C16orf5 gene

The gene C16*orf*5 consists of an open reading frame of 786 bp, encoding a putative protein of 261 aminoacids and a large 1517 bp 3'-untranslated region (Figure 3.4). The putative translation initiation site begins with an ATG at the nucleotide 298, and the sequence surrounding the site conforms to Kozak's rule (Kozak 1991, 1996). Two polyadenylation sites are present, one at 2144bp and the second at 2756 bp (Figure 3.4).

The partial genomic structure of the gene was determined using the PAC 102J11 by the bubble PCR approach (Gecz et al., 1997) using gene specific primers and bubble linker-specific primers. This procedure allowed the identification of at least four tentative exonic boundaries in the C16*orf*5 gene. Later in the course of this study, the identification of BAC 351P16 sequence spanning the C16*orf*5 gene enabled the determination of intron–exon organisation of C16*orf*5 gene, which is presented in Chapter 7.

3.3.6 Computational analysis of the C16orf5 cDNA sequence

Database searching using the entire C16*orf*5 sequence showed no significant homology to known proteins and domains. Searching of the PROSITE database with C16*orf*5 protein sequence predicted a single N-glycosylation site,

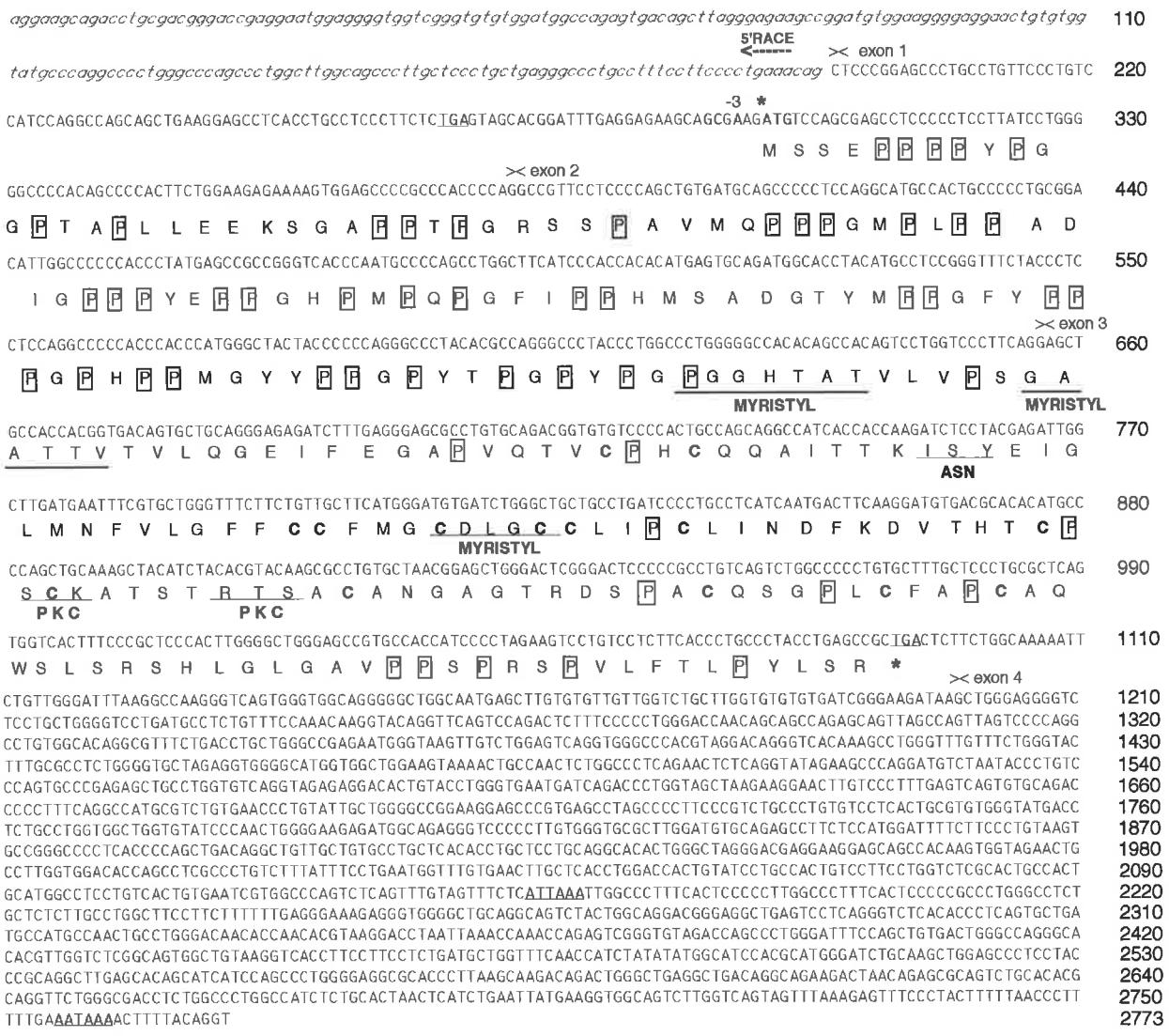


Figure 3.4 Sequence of C16orf5. Nucleotide sequence of the C16orf5 cDNA and the deduced amino acid sequence of its predicted protein are shown. Asterisks denote the start and termination codons. Bold capital letters in the sequence flanking the start codon, (**GCGAAGATGT**) conforms to the optimum Kozak consensus. The first 193 nucleotides were obtained by 5'RACE. The proposed polyadenylation signals (Sheets et al. 1990; Wahle and Keller 1992), and an in frame stop codon before the initiation codon are underlined. Proline residues are boxed and cysteine residues are shown in boldface.The predicted 1N-glycosylation (ASN), 3N-myristoylation (MYRISTYL) and 2 protein kinase C (PKC)phosphorylation sites are indicated. Each exon boundary is shown by >< above the nucleotide sequence. The nucleotide sequence reported here has been deposited in GenBank under the accession number AF131218.

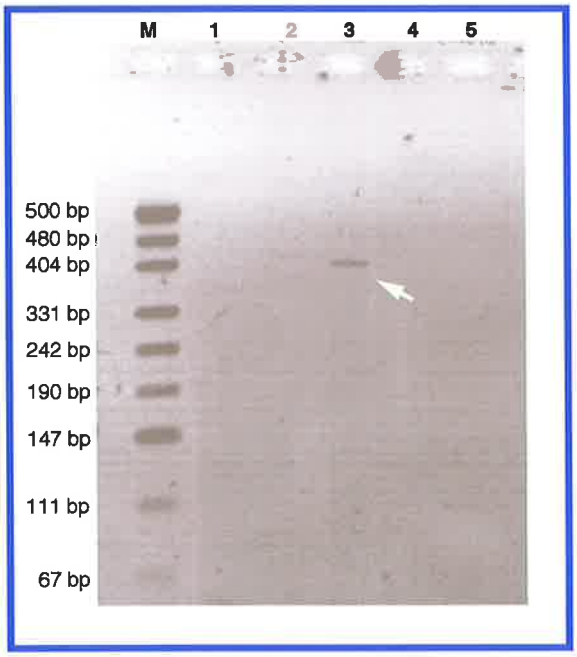


Figure 3.5 Agarose gel analysis of 5' RACE products of C16*orf*5 transcript. Lane1, First PCR amplification of dC tailed cDNA using the Abriged Anchor primer (AUAP) supplied with 5' RACE kit and GSP2 primer. No PCR product was amplified after first PCR amplification. Lane 3, Reamplification of the primary 5'RACE product using AUAP and nested GSP3 primer. Arrow denotes the 5'RACE product of approximately 415 bp in size. Lane 5, represents no DNA control. No sample was loaded in the lanes 2 and 4. Lane M, pUC 19/*Hpa* II marker DNA. 15 ul of each PCR reactions (lanes 1 and 3) were analyzed on 2.5% agarose gel in 1X TBE stained with 0.5 ug/ml ethidium bromide. Fetal brain poly A+ RNA for 5' RACE analysis was kindly provided by Dr Jozef Gecz, Department of Cytogenetics, WCH, Adelaide.

three N-myristoylation sites and two protein kinase C phosphorylation sites (Figure 3.4). Computer prediction of the subcellular localisation using PSORT II indicated that C16*orf*5 would most probably localize to the nucleus. C16*orf*5 sequence has an unusually high content of proline residues (40% over 104 residues) at the N-terminus of the protein. The C-terminus of the protein is also cysteine rich with 14 cysteine residues present (Figure 3.4).

3.3.7 Expression analysis of C16orf5 gene

The expression profile of C16*orf*5 was investigated using a Multiple Tissue Northern Blot and RNA dot blot. The Multiple Tissue Northern Blot contained poly (A^+) mRNA from various adult human tissues, including heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas. The Northern blot when probed with the insert of cDNA clone AA070715 predicted a single band of approximately 2.4 kb in size with the strongest signal in brain but also expression in heart, skeletal muscle, kidney, pancreas and liver (Figure 3.6). This is likely to correspond to transcripts that are generated using a polyadenylation site starting at 2,144 bp (Figure 3.4). Some transcripts in the NCBI database dbEST were generated using the second polyadenylation site starting at 2,756 bp (Figure 3.4), however there was no clear evidence of transcripts of this size from the Northern blot. Also probed was a RNA dot blot that contained poly A^+ RNA from 50 different human tissues with the quantity of RNA normalised with respect to several housekeeping genes. This procedure showed that that there was expression in all the brain samples (whole brain, amygdala, caudate nucleus, cerebellum, cerebral cortex, frontal lobe, hippocampus, medulla oblongata, occipital lobe, putamen, substantia nigra, temporal lobe, thalamus, subthalamic nucleus) and

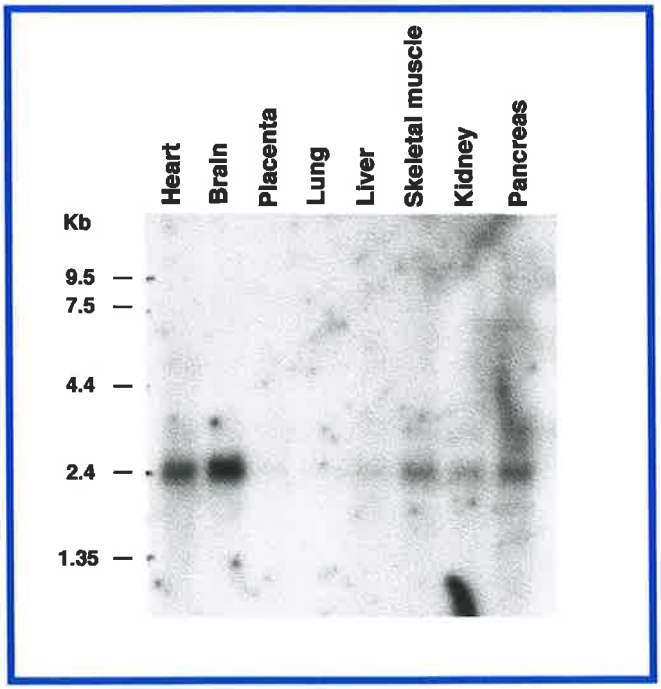


Figure 3.6 Detection of C16orf5 transcript by Northern blot analysis. The filter with 2 ug of poly A+ RNA from different tissues in each lane was probed with the insert of cDNA AA070715. A single transcript of 2.4 kb was detected. The position of the size markers is indicated.

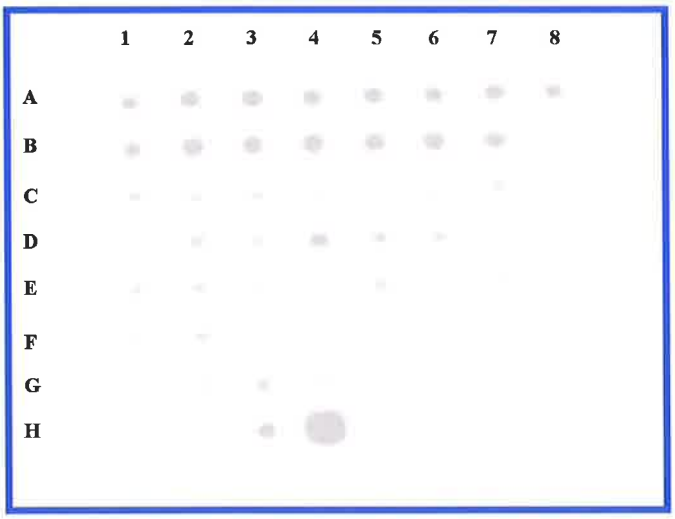


Figure 3.7 Human RNA dot blot expression analysis of C16*orf*5 gene. The RNA dot blot was hybridized with the insert of cDNA clone AA070715. The filter contained 100-500 ng/1-mm dot poly A+ RNA from 50 different tissues. Highest level of expression was obtained in all the tissues of the brain (A1, whole brain; A2, amygdala; A3, caudate nucleus; A4, cerebellum; A5, cerebral cortex; A6, frontal lobe; A7, hippocampus; A8, medulla oblongata; B1, occipital lobe; B2, putamen; B3, substantia nigra; B4, temporal lobe; B5, thalamus; B6, subthalamic nucleus) and (B7) spinal cord. Heart (1C) and all other tissues (Figure 3.6) showed weak expression. The membrane also contained negative controls that are represented by dots: H1) yeast total RNA, H2) yeast tRNA, H5) poly r(A), H6) human C_0t 1 DNA. These negative controls did not hybridise to the gene specific probe. Hybridisation of the dots H3 and H4, which contained E *coli* RNA and DNA respectively, is likely due to the presence of E *coli* DNA in the DNA prepared for the cDNA clone AA070715. D4 represents pituitary gland.

spinal cord, while heart and all other tissues showed very weak or absent expression (Figure 3.7).

3.3.8 Identification of DNA clones spanning the C16orf5 gene

In an attempt to determine the relative position of C16*orf*5 gene and the 16p13.3 breakpoints of the two patients with *de novo* balanced translocations, the cDNA probes AA070715 and H07909 were used to screen two different PAC libraries (section 3.2.7). The insert sizes of DNA clones were determined and restriction mapped using pulsed-field gel electrophoresis (PFGE) (section 2.3.4.4.3) and the infrequently cutting restriction enzymes *Not*I and *Eag*I. This method allowed the determination of clone overlaps and the clones were arranged into a possible contig. A further walk was undertaken from one of the furthest extending PACs (DNA clone 43I06) by screening a BAC library (section 3.2.7) with the SP6 end probe generated from the insert. An example of restriction fragment pattern obtained for PAC and BAC DNA clones screened using cDNA clones AA070715 and H07909 and SP6 end probe generated from PAC 43I06 is shown in Figure 3.8 A and B. The DNA restriction fragments were resolved within resolution limits ranging from 10 kb-200 kb (Figure 3.8 A). To further resolve the high molecular weight restriction fragments the DNA resolution was also carried out within limits ranging from 5kb-50kb (Figure 3.8 B). Subsequently, the PFGE filter containing the PAC and BAC DNA clones (Figure 3.8 B) was prepared. The contig spanning C16*orf*5 gene was confirmed by Southern analysis using the SP6 end probe generated from PAC 43I06 (Figure 3.9 A). The contig was also probed with cDNA clones AA070715 and H07909 to identify the *Not*I/*Eag*I restriction fragment containing the C16*orf*5 gene. The gene was contained in the 50 kb fragment of this contig.

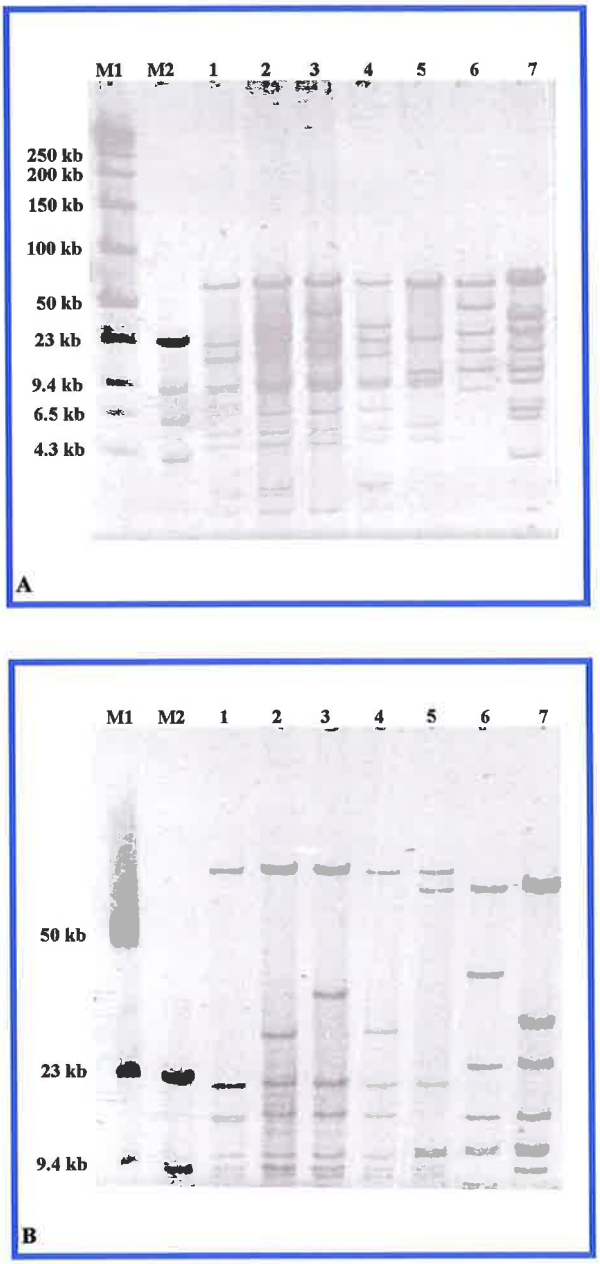


Figure 3.8 An example of an agarose gel showing restriction fragment pattern for PAC and BAC DNA clones spanning the C16*orf*5 transcript. Samples: M1) Low molecular weight PFGE marker, M2) Drigest, 1) BAC 709D24, 2) BAC 707M17, 3) 734B17, 4) 793I17, 5) 834A11, 6) PAC 102J11, 7) PAC 43I06. The conditions for resolution of DNA fragments using PFGE were: gel: 1% agarose gel in 0.5 X TBE, running temperature 14^{o}C, angle of electrodes:120^{o}.

(A) The DNA fragments in gel A were resolved within the limits ranging from 10 kb-200 kb. The gel run time was 20 hours. **(B)** The DNA fragments in gel B were resolved within the limits ranging from 5 kb-50 kb for the resolution of higher molecular weight DNA fragments. The gel run time was 20 hours

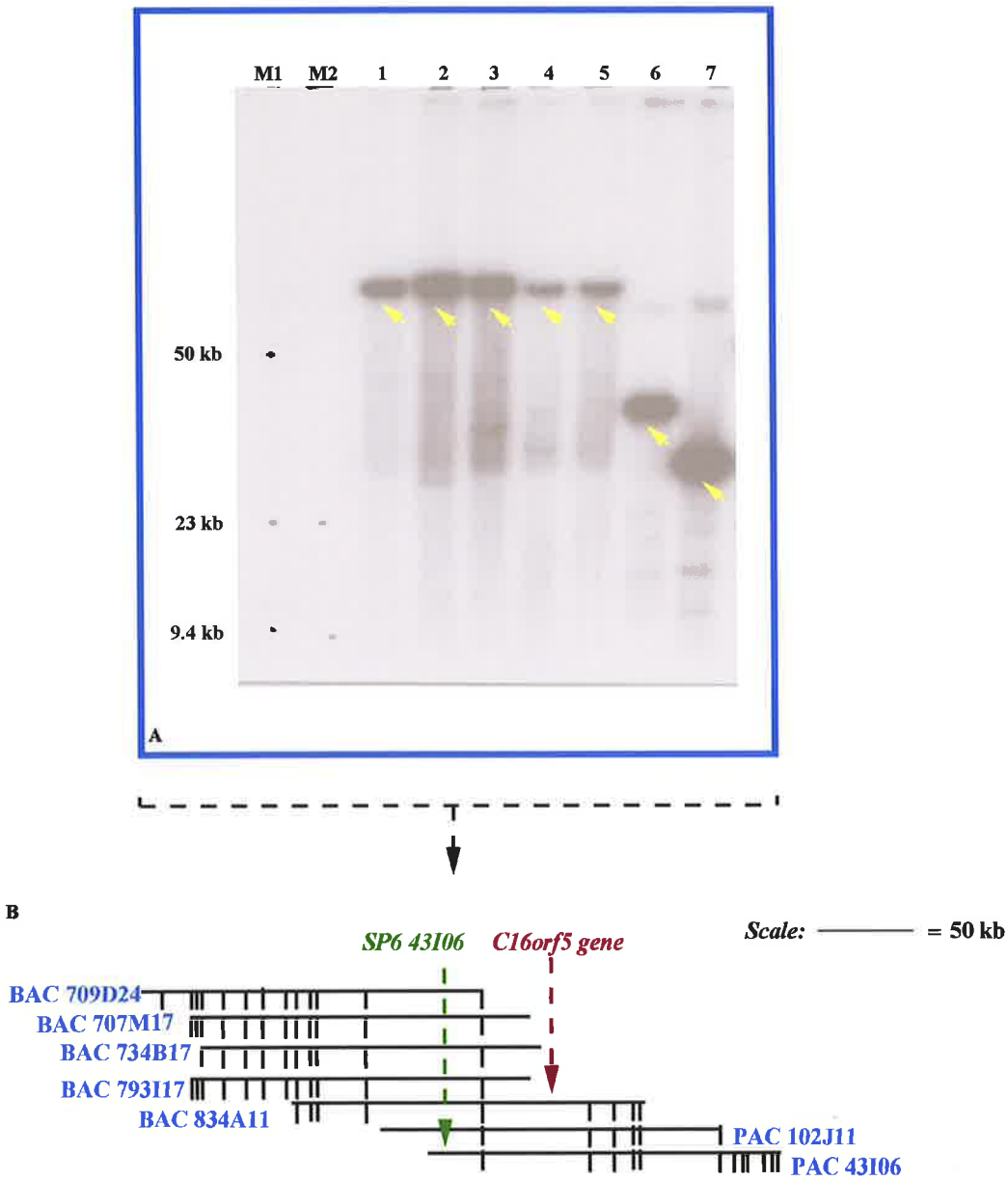


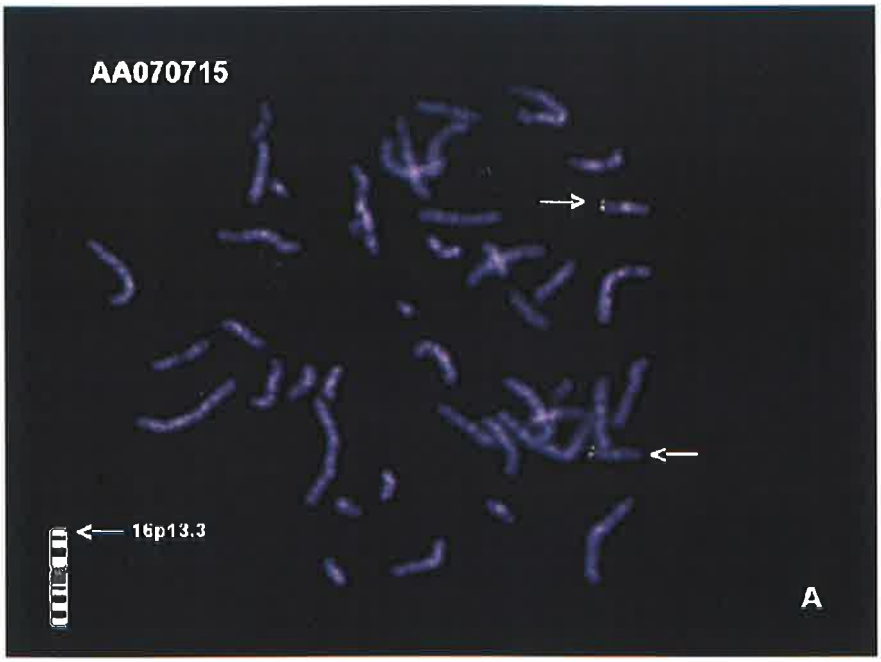
Figure 3.9 Southern hybridisation of PFGE filter containing PAC and BAC DNA clones spanning the C16*orf*5 gene with SP6 end probe generated from the PAC 43I06. **(A)** Samples: M1) Low molecular weight PFGE marker, M2) Drigest, 1) BAC 709D24, 2) BAC 707M17, 3) 734B17, 4) 793I17, 5) 834A11, 6) PAC 102J11, 7) PAC 43I06. The filter was hybridised as mentioned in section 2.3.4.8 with radiolabelled SP6 43I06 end probe. Autoradiography was carried out at -70^{o}C for overnight to visualise the hybridised bands. The arrows depicts the fragments positive for SP6 43I06 end probe. **(B)** Arrangement of the clones spanning the C16*orf*5 gene in a contig. The positions of SP6 43I06 end probe and C16*orf*5 gene on this contig are indicated. The *Not*I and *Eag*I sites are indicated by horizontal lines.

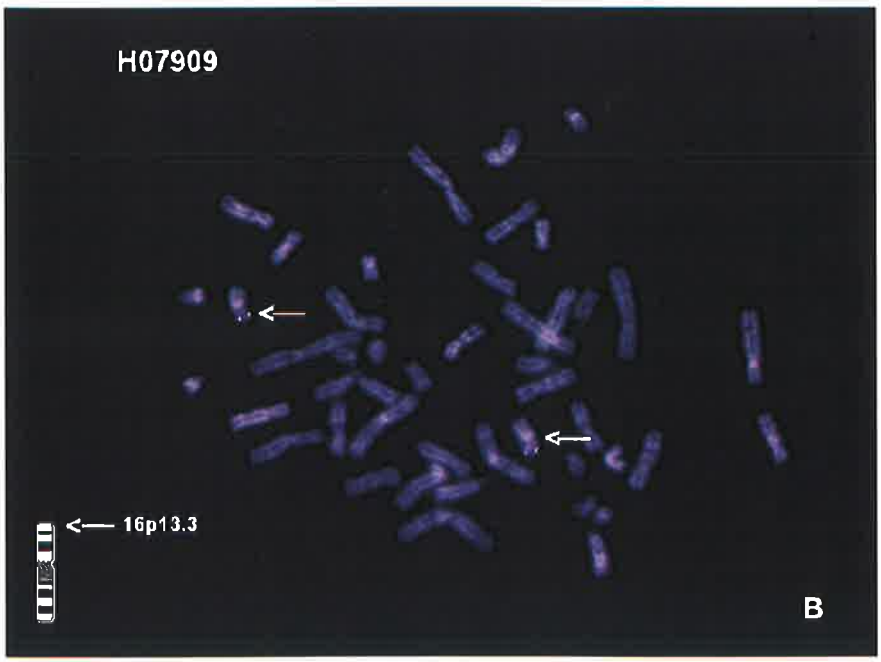
The contig spanning the gene and the position of C16*orf*5 gene on this constructed contig is depicted in (Figure 3.9 B).

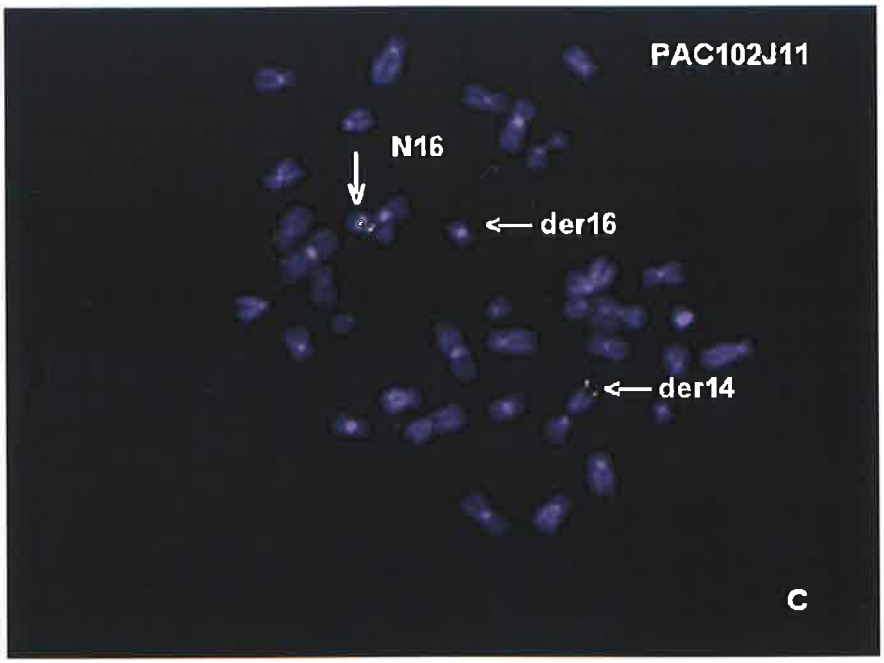
3.3.9 Fluorescent in situ hybridization

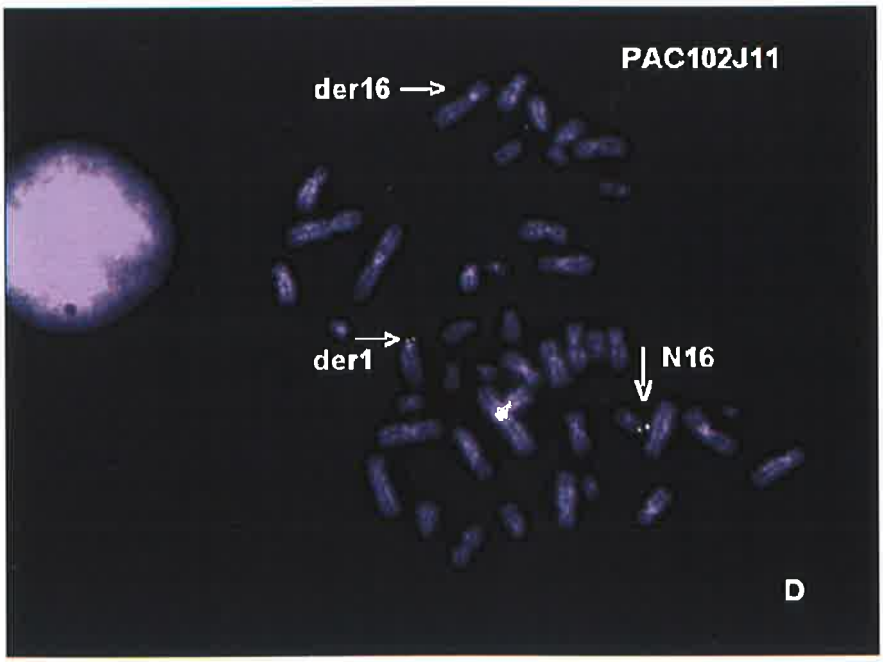
Chromosomal assignment of the C16*orf*5 gene was performed by FISH analysis of metaphase chromosomes from a normal male. Twenty metaphases all showed signal on one or both chromatids of chromosome 16 at 16p13.3 for the cDNA probes AA070715 and H07909. 3 to 5 non-specific dots were observed in these 20 metaphases respectively (Figure 3.10 A and B). In order to determine the relative position of C16*orf*5 gene with respect to the 16p13.3 breakpoints, FISH analysis was done with the PAC DNA clone 102J11 spanning the C16*orf*5 transcript to the metaphase chromosomes of the two patients with t(14,16) and t(1;16) *de novo* translocations. The PAC DNA clone 102J11 showed distal signal to the 16p13.3-translocation breakpoints in both the patients (Figure 3.10 C and D).

Figure 3.10 Fluorescence in situ hybridisation. **(A and B)** Metaphase showing FISH with cDNA clones AA070715 and H07909 respectively. Normal male chromosomes stained with DAPI. Hybridisation sites on chromosome 16 are indicated by arrows. **(C)** FISH analysis of the PAC DNA clone 102J11 spanning the C16*orf*5 transcript to the metaphase chromosomes from the patient having t(14;16) *de novo* balanced translocation. The FISH signals were seen on the normal chromosome 16 and derivative chromosome 14. **(D)** FISH analysis of the PAC DNA clone 102J11 to the metaphase chromosomes from the patient having t(1;16) *de novo* balanced translocation. The FISH signals were seen on the normal chromosome 16 and derivative chromosome 1. This indicated that the PAC 102J11 was distal to the 16p13.3 translocation breakpoints of t(14;16) and t(1;16).


AA070715
16p13.3
A


H07909
16p13.3
B

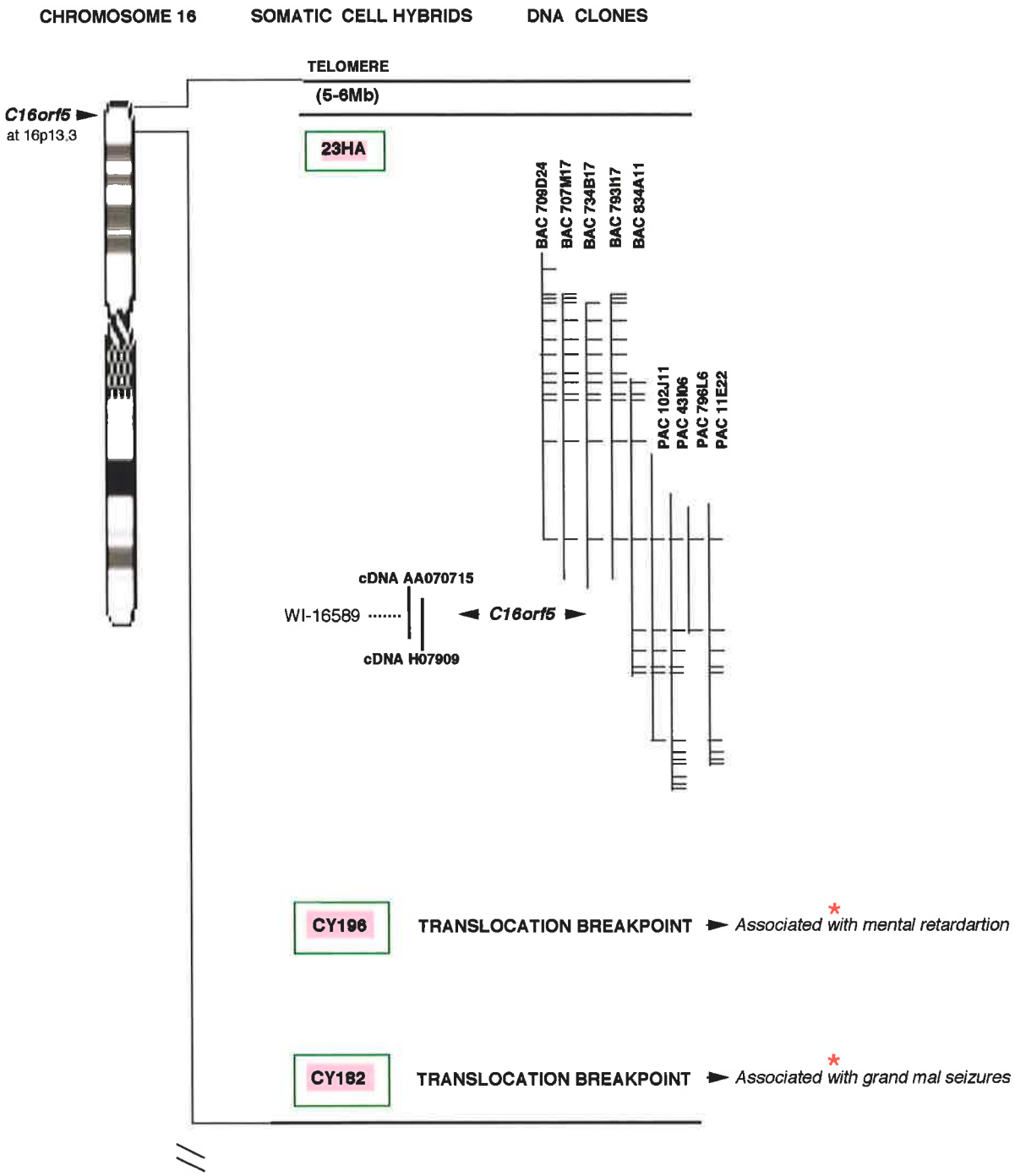

PAC102J11
N16
der16
der14
C


PAC102J11
der16
der1
N16
D

3.4 Discussion

The DNA probes were localised to the short arm of chromosome 16 in the vicinity of the hybrid breakpoints CY196 and CY182 by PCR analysis of the somatic cell hybrids at 16p13.3 (Callen et al., 1995). These mapping data separated the chromosome breakpoints contained in the hybrids CY196 and CY197 (Table 3.2 and Figure 3.1). There were no markers identified earlier separating these two breakpoints in the previously reported map of chromosome 16 (Callen et al., 1995). The order of the hybrid breakpoints CY177, CY182-CY196, CY197 reported in Callen et al., (1995), was refined to pter-CY196-CY197, CY182, CY177. Later in the course of study additional markers were identified in the region that separated the chromosome breakpoints in the hybrids CY197, CY182 and CY177, which are presented in Chapters 4 and 6 respectively. The preliminary physical mapping efforts in the vicinity of the hybrid interval defined by CY196 and CY182 identified a novel proline rich gene C16*orf*5. The PCR analysis of somatic cell hybrids at 16p13.3 localised the C16*orf*5 gene in the adjacent hybrid interval to the hybrid breakpoints CY196 and CY182. The localisation of this gene was also confirmed by Southern hybridisation using cDNA clones AA070715 and H07909. Schematic representation of C16*orf*5 gene with respect to the hybrid breakpoints CY196 and CY182 is presented in Figure 3.11.

A notable feature of the peptide includes its proline richness (40% over 104 residues) at the N-terminus of the predicted protein. Proline richness is a characteristic of DNA binding proteins (Bennett and Reed, 1993). Proline rich activation domains are associated with transcription factors like CTF/NF-1, AP-2,



*The clinical case reports for the two patients with t(1;16) and t(14;16) de novo balanced translocations are presented in Chapters 5 and 6 respectively.

Figure 3.11 Schematic representation of the position of C16orf5 in human chromosome 16p13.3. Somatic cell hybrid analysis localized the gene to chromosome 16p13.3 in the hybrid interval defined by 23HA and CY196-CY182. The position of the gene is indicated by arrows. The NotI and EagI restriction sites are indicated by horizontal lines. The PAC and BAC contig spanning the gene is also shown. The *Not*I and *Eag*I sites are indicated by horizontal lines.

and steroid receptors for progesterone and estrogen (Mermod et al., 1989). The predicted C16*orf*5 protein is also cysteine rich in the C-terminal half of the gene, with 14 cysteine residues present that may participate in intra-and/or interchain disulfide bonding (Marquardt et al., 1998). The C16*orf*5 coding sequence also contains a single N-glycosylation site, three N-myristoylation sites and two protein kinase C phosphorylation sites, which suggests that the protein might encode a secreted or plasma membrane protein (Habuchi et al., 1998).

The C16*orf*5 gene is a 2.4-kb transcript and is highly expressed in the brain. The expression patterns determined from the normalised RNA of the dot blot show high levels of expression in all the tissues of the brain. Although the function of the C16*orf*5 protein is unknown, its proline richness, presence of cysteine residues, predicted nuclear localisation, and an abundant brain specific expression suggest that it is likely to have an important role in the development and function of neuronal cells.

Because this gene was highly expressed in the brain and was localized within the hybrid interval 23HA and CY196-CY182, it was a candidate that it may be disrupted by the two translocations t(1;16) and t(14;16)at 16p13.3 involved in mental retardation and epilepsy respectively. The most likely possibility was that this transcript is disrupted by the 16p13.3 breakpoint of t(1;16), as this gene was localized to the immediate distal hybrid interval to the chromosome breakpoint in CY196. Therefore, in an attempt to determine the relative position of the C16*orf*5 gene with respect to the 16p13.3 breakpoints, PAC and BAC DNA libraries were screened to identify the DNA clones spanning the gene. Subsequently, FISH analysis was carried out on the metaphase chromosomes of the two patients using

PAC DNA clone 102J11, which contained the entire C16*orf*5 gene. The FISH data for PAC DNA clone 102J11 placed the gene distal to the 16p13.3 translocation breakpoints of the two patients with t(1;16) and t(14;16) *de novo* balanced translocations. However, if only a small fragment of the PAC is proximal to the breakpoints it is possible to obtain a distal FISH signal on the metaphase chromosomes because the major length of the probe lies distal to the breakpoint. Therefore, in order to determine if the C16*orf*5 gene was directly disrupted by the 16p13.3 breakpoints, it was necessary to generate the 5' and 3'end probes from the PAC DNA clone 102J11 and to map them with respect to the CY196 and CY182 hybrid breakpoints, which contained the der(16) of the two translocations. It was then decided to construct a long range YAC restriction map of the breakpoint region at 16p13.3 with the aims to first localise the 16p13.3 breakpoints, and then to estimate if the C16*orf*5 gene was directly disrupted by the 16p13.3 breakpoints involved in mental retardation and severe seizure. The YAC long-range restriction map of the 16p13.3 breakpoints region is presented in Chapter 4.

Long Range YAC Pulsed-field Map of the Breakpoints Region At 16p13.3

Chapter 4

4.1 Introduction

The Yeast artificial chromosome (YAC) cloning system, which is capable of incorporating large fragments (greater than one megabase) of exogenous DNA in a single clone (Burke et al., 1987), facilitates the construction of long-range physical maps. Long range YAC physical maps of the entire genome have previously been established by Cohen et al., (1993), Chumakov et al., (1995), Hudson et al., (1995) and specifically of chromosome 16 by Doggett et al., (1995). YAC maps have made important contributions to the identification of many disease genes. An example is that of Huntington disease, where a YAC map spanning over a megabase at chromosome 4p16.3 between DNA markers D4S95 to D4S127 was highly beneficial in the identification of candidate genes in the region (Rommens et al., 1993). Geraghty et al., (1993), also used YACs mapping in the OATL1 (ornithine -δ-aminotransferase gene) region at Xp11.2 associated with retinal disorders, for the identification of candidate transcripts.

Even with the advent of BAC maps, YACs can still provide a useful method for constructing long-range restriction maps. The large DNA inserts of YACs can be resolved using pulsed-field gel electrophoresis (PFGE) (Chu et al., 1986). This technique is based on the principles of contour-clamped homogeneous electric field (CHEF) that alternates between two orientations that are 120° apart. The electric field is generated by a method in which multiple electrodes are arranged along a polygonal contour and clamped to predetermined electric potentials (Chu et al., 1986).

In the present study, the YAC pulsed-field restriction map of the breakpoints region at 16p13.3 was constructed with the major aim to localise the two breakpoints of the t(1;16) and t(14;16) contained in CY196 and CY182 respectively. This was necessary to determine if the C16*orf*5 gene is directly disrupted by either of these breakpoints. This was approached by generating probes from the 5' and 3' ends of the insert of PAC clone 102J11, which was placed distal to the 16p13.3 breakpoints of t(1;16) and t(14;16) by FISH (section 3.3.9), and mapping these end probes with respect to the pulsed-field restriction map of YAC in the region. The YAC map will allow integration of BAC and PAC clones to enable the cloning of the 16p13.3 breakpoints contained in CY196 and CY182.

In this Chapter the YAC PFGE restriction map of the breakpoints region at 16p13.3 and mapping of the C16*orf*5 gene with respect to this YAC map of the region will be presented.

4.2 Methods

The methods specific for this Chapter are described in detail. Other general techniques used in this study are presented in Chapter 2.

4.2.1 Preparation of agarose blocks containing yeast DNA

The YAC DNA was prepared in agarose blocks using the lithium lysis method described by Chaplin and Brownstein, (1995). The YAC cultures were streaked on AHC agar (section 2.2.6.2) plates, which were incubated at 30°C for 48 hours to obtain a single isolated colony. The single yeast colony was then inoculated in 200 ml

AHC broth media (section 2.2.6.1) containing 50 μg/ml of tetracycline at 30°C with shaking for 48 hours. After incubation, the yeast cells were pelleted at 3000 rpm for 10 minutes at 4°C. The pellets were resuspended in 10 ml of 50 mM EDTA solution, then pelleted again as before. The supernatant was decanted and the yeast cell pellet was resuspended in 1 ml of a solution containing 1 M sorbitol, 20 mm EDTA, 14 mM 2-mercaptoethanol and 1 mg/ml lyticase. To this was added 2 ml (an equal volume to the amount of resuspended pellet) of molten 2% low melting point agarose in 1 M sorbitol, 20 mM EDTA and 14 mM 2-mercaptoethanol at 50°C (agarose was dissolved by boiling, and then cooled to 50°C in waterbath before the addition of 2-mercaptoethanol). The resulting solution was dispersed into plug molds, which were set at 4°C for 15-20 minutes. Plugs were extruded into 20 ml of 1 M sorbitol, 20 mM EDTA, 14 mM 2-mercaptoethanol, 10 mM Tris-HCl (pH 7.2), 1 mg/ml lyticase solution, then incubated at 37°C for 1 hr. After incubation, the solution was removed and replaced with 25 ml of lithium lysis solution (1% lithium dodecyl sulphate, 100 mM EDTA, 10 mM Tris-HCl, pH 7.2), that is filtered and prewarmed at 50°C. The resuspended plugs were incubated at 37°C for 1 hour, and then the solution was drawn off and replaced with the fresh lithium lysis solution following an overnight incubation at 37°C. After overnight incubation, the lithium lysis solution was removed, and the blocks were washed twice in 50 ml of TE, first at 50°C for 60 minutes, then 50°C for 30 minutes. After washing the blocks were stored in TE at 4°C.

4.2.2 Restriction digestion of YAC DNA in agarose block

(Modification of Finney, 1994)

The agarose block containing the YAC DNA was first washed with 100 µl of 10X restriction buffer (100 mM Tris-HCl, pH 7.2; 100 mM $MgCl_2$; 10 mM DTT and 50 mM NaCl) for 1 hr at room temperature. The YAC DNA was digested in the presence of 10 µl of 10X restriction buffer, 1 µl of 10 mg/ml BSA and 20 units of restriction enzyme (New England Biolabs) in a final volume of 100 µl. The YAC DNA was digested using following combinations of enzymes: *Eag*I + *Not*I; *Not*I + *Asc*I; *Asc*I + *Bss*HII; *Asc*I + *Mlu*I; *Not*I + *Mlu*I; *Bss*HII + *Not*I; and *Asc*I + *Eag*I. The reaction was incubated at 37°C overnight.

4.2.3 Pulsed-field gel electrophoresis (PFGE)

Agarose blocks containing immobilised restricted YAC DNA were loaded into wells prior to gel submersion in the PFGE apparatus buffer tank. Each block was cut to the size of a well, then loaded using a small spatula, taking care that the interface between the block and the front of the well contained no bubbles. The top of each well was then sealed with 1% agarose gel before submersion. PFGE (2.3.4.4.3) was achieved using a BioRad CHEF Mapper™ apparatus. This apparatus has an integrated software system whereby the parameters defining the required window of DNA size separation are entered. An algorithm automatically selects the appropriate switching ratios to maximise resolution within that window. Alternatively, the software included barcode programs for the resolution of specific size ranges of DNA molecules.

4.2.4 Preparation of YAC pulsed-field gel filters

The YAC DNA was digested using the restriction enzymes mentioned in section 4.2.2. DNA was separated by PFGE (2.3.4.4.3) and subsequently the DNA was transferred to Hybond N+ ^{TM} (Amersham) membranes as per the method described in section 2.3.4.7.1. Gels were UV nicked using BioRad Gene Linker^{TM} UV Chamber (CA, USA), before the transfer of DNA.

4.2.5 DNA probes

The YAC PFGE filters were probed with left and right vector arm probes, total human BAC DNA probes (250-300 ng/μl) and C16*orf*5 gene specific probes. The undigested pUC DNA and pBR322 PCR product (515 bp) generated with the primer pairs (Forward, 5'-AGG TGC GGT TGC TGG CGC CTA TAT C–3'; Reverse primer, 5'-ATG CCG GCG ATA ATG GCC TGC TTC T-3') were used as the left and right vector arm specific probes respectively. The restriction map was constructed manually.

4.2.6 Labelling of DNA probes

The DNA probes were purified using the methods described in section 2.3.2. Subsequently, the probes were labelled with α ^{32}P as described in section 2.3.4.5.

4.2.7 Fluorescent in situ hybridisation (FISH) analysis

The FISH analysis of DNA clones, YAC 761C2, and BACs 375G12, 315L9 and 118C2 to the metaphase chromosomes of the two patients with *de novo* balanced

translocations t(1;16) and t(14;16) was performed as described in section 2.3.11 by Helen Eyre (Department of Cytogenetics and Molecular Genetics, WCH, Adelaide).

4.2.8 PCR mapping of DNA probes

The DNA probes localised in the vicinity of the 16p13.3 breakpoints of the t(1;16) and t(14;16) were integrated into the YAC restriction map of the breakpoints region by PCR analysis of the YAC 761C2 and BACs 375G12, 315L9 and 118C2. An agarose block containing the YAC DNA was dissolved in 100 µl of TE and denatured at 100°C for 10 minutes. 1 µl of this suspension was used for PCR amplification of the YAC DNA in 10 µl PCR reaction. The amount of BAC DNA used in 10 µl PCR reaction was 50 ng. The PCR reaction conditions used were as described in section 2.3.5. The primers for DNA probes (Table 4.2) incorporated onto the restriction map of the YAC 761C2 were designed as mentioned in section 2.3.5.1.

4.3 Results

4.3.1 Identification of YAC clones: mY761C2, mY929E9 and mY665G1

The database searches of the Whitehead Institute (http://carbon.wi.mit.edu:8000/cgi-bin/contig/yac-info) using the DNA markers D16S423 (AFM249yc5), AFM339xg1 and AFM288zf9 localised in the vicinity of the 16p13.3 breakpoints (Table 3.2), identified a YAC contig spanning the hybrid interval defined by CY196 and CY182. Three overlapping clones (YACs 761C2 929E9 and 665G1) from this contig were

expected to span the breakpoints region and were therefore obtained for the construction of the long-range pulsed-field YAC map of the region.

4.3.2 Estimation of the YACs sizes by PFGE

In order to determine the sizes of the respective YACs, the YAC DNA was prepared in agarose blocks (4.2.1) and electrophoresed uncut on a pulsed-field gel (section, 2.3.4.4.3) followed by the preparation of YAC PFGE filters (4.2.4). The filters were probed with left and right vector arm probes pUC19 and pBR322 PCR product respectively. The size of YAC 761C2 was estimated to be approximately 850 kb (Figures, 4.1 A and B), which was same as the size mentioned for this YAC in the Whitehead database. The band corresponding to the YAC 929E9 on the autoradiographs in the Figures 4.1 A and B was approximately 600 kb. However, the size mentioned for this YAC in the Whitehead database was 1640 kb. This suggested that the YAC 929E9 isolate studied was deleted. For the YAC 665G1, no specific YAC DNA band was detected on the autoradiographs in the Figures 4.1 A and B. It was likely that this YAC culture stab received was contaminated. The YACs 929E9 and 665G1 were therefore discarded from further Southern analysis.

4.3.3 BAC DNA clones 375G12, 315L9 and 118C2

The BAC DNA clones 375G12, 315L9 and 118C2 (RPCI-11 human BAC library) were identified from the BAC contig at 16p13.3 constructed by the Los Alamos National laboratory, New Mexico, using the DNA markers localised in the vicinity of the 16p13.3 breakpoints contained in CY196 and CY182 (Table 3.2). These BACs were kindly provided by Dr Norman Doggett, Los Alamos National laboratory.

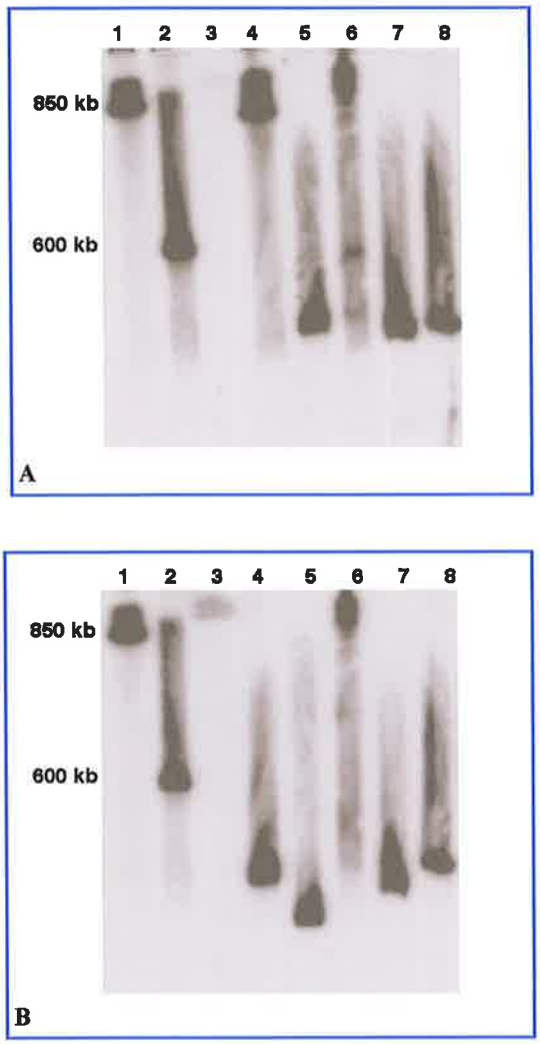


Figure 4.1 Autoradiographs of YAC PFGE filters probed with left and right vector arm probes pUC19 and pBR322 respectively. Samples:1)YAC 761C2, uncut; 2)YAC 929E9, uncut; 3) YAC 665G1, uncut. Lanes 4-8 corresponds to YAC 761C2 digested with restriction enzymes *Asc*I, *Bss*HII, *Not*I, *Eag*I, and *Mlu*I respectively. PFGE was performed using BioRad CHEF Mapper apparatus. Conditions used for resolution of DNA fragments within size range 200-2,200 Kb were: gel 1% agarose in 0.5X TBE, running temperature: 14^{O}C, angles of electrodes: 120^{O}, voltage: 6V/cm, run time: 15 hours. Protocol was obtained from CHEF Mapper instruction manual (Barcode # 10). Sizes of YACs are indicated.

(A) Autoradiograph shown in 4.1 A was probed with LHS YAC probe (pUC19).

(B) Autoradiograph shown in 4.1 B was probed with RHS YAC probe (pBR322).

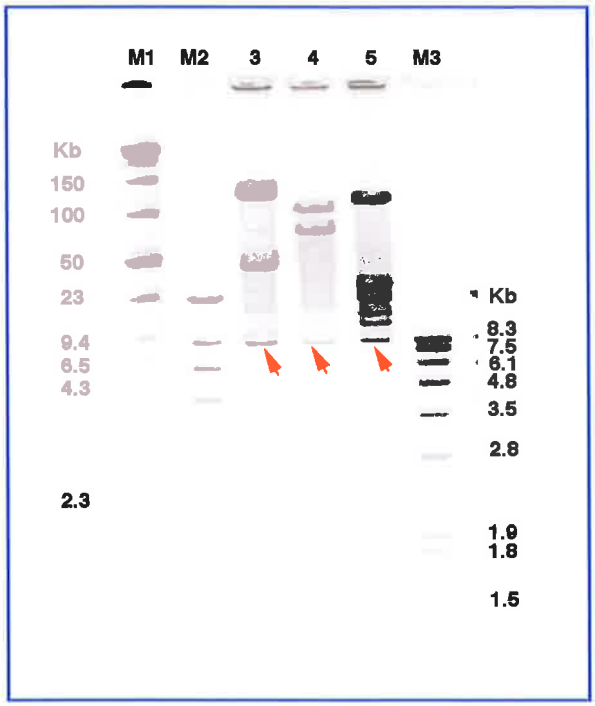


Figure 4.2 Estimation of the sizes for BACs 375G12, 315L9 and 118C2 by PFGE. Samples: M1) Low range pulsed-field gel marker; M2) molecular weight marker, Drigest; 3) BAC 375G12; 4) BAC 315L9; 5) BAC 118C2; M3) molecular weight marker, SPP1. These BACs were kindly provided by Dr. Norman Doggett (Los Alamos National Laboratory). The BAC DNA was digested using restriction enzymes *Not*I and *Eag*I. Subsequently, the DNA was resolved by PFGE. The conditions for resolution of DNA fragments within the range 5 kb - 150 kb were: gel: 1% agarose gel in 0.5X TBE, running temperature: 14^{o}C, angles of electrodes: 120^{o}, run time: 15 hours. The restriction fragments obtained for respective clones were added leaving the vector bands to estimate the sizes of respective BAC DNA clones. The red arrows depicts the vector band of the BAC vector pBACe3.6. The sizes of BACs as estimated by PFGE were: BAC 375G12, 162 kb; BAC 315L9, 190 kb; and BAC 118C2, 185 kb. The position and sizes of marker bands are indicated.

The BAC DNA was prepared (2.3.1.3) and digested using the restriction enzymes *Eag*I and *Not*I and the sizes were estimated by PFGE (2.3.4.4.3). The sizes of BACs 375G12, 315L9 and 118C2 were found to be 162 kb, 190 kb and 185 kb respectively (Figure 4.2). The YAC 761C2 PFGE filters were hybridised with these BACs in order to determine the YAC restriction fragments to which they were contained.

4.3.4 Characterisation of YAC 761C2 by PFGE

The size of the YAC 761C2 (850 kb) as estimated by PFGE (Figures 4.1 A and B) corresponded to the size mentioned in the Whitehead database, therefore this YAC was used in further mapping experiments. In order to construct the YAC 761C2 restriction map, the DNA in agarose block was digested (section 4.2.2) and resolved using PFGE, at two different resolution limits, 50-500 kb and 250 kb-1.5 Mb. Subsequently the YAC PFGE filters were prepared and probed with vector arm probes (pUC19 and pBR322) and BAC DNA probes.

The YAC 761C2 filters were first probed with the α ^{32}P labelled left and right vector arm probes pUC19 and pBR322 respectively to identify the restriction fragments located at both the ends of human insert of YAC 761C2. The examples of banding pattern are depicted in Figures 4.3 A and B. The radioactivity associated with pUC19 and pBR322 probes was stripped (section 2.3.4.9) from the filters and the filters were reprobed with BAC DNA probes BACs 375G12, 315L9 and 118C2. An example of the banding pattern obtained with the BAC 118C2 is shown in Figure 4.3 C. This enabled the localisation of BACs on the YAC PFGE map.

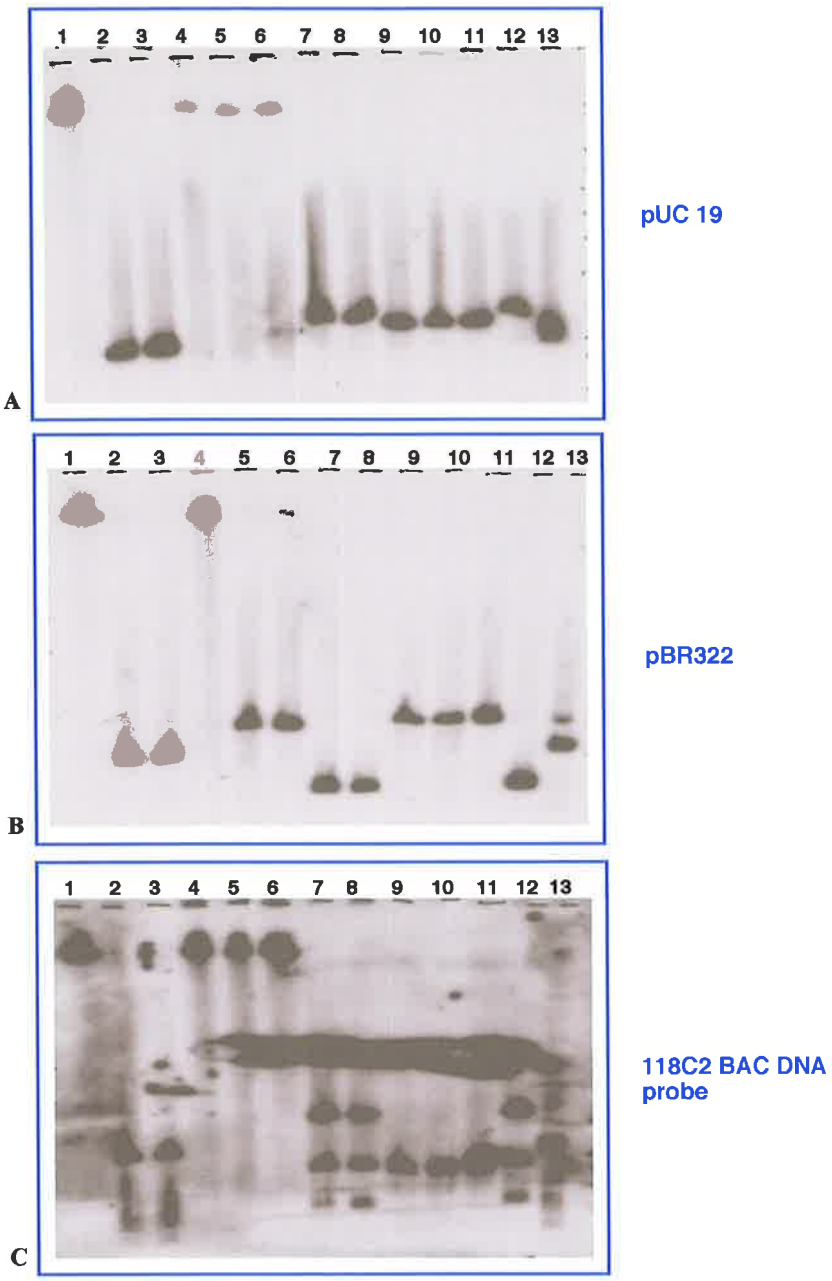


Figure 4.3 Autoradiographs showing examples of the hybridization pattern of YAC 761C2 to vector arm probes pUC 19 and pBR322 and a human BAC DNA probe. Samples: 1) YAC 761C2, uncut. Lanes 2-13 represents YAC 761C2 DNA digested with following restriction enzymes: 2) *EagI*, 3) *Eag*I and *Not*I 4) *Not*I, 5) *Not*I and *Asc*I, 6) *Asc*I, 7) *Asc*I and *Bss*HII, 8) *Bss*HII, 9) *Asc*I and *Mlu*I, 10) *Not*I and *Mlu*I, 11) *Mlu*I, 12) *BssH*II and *Not*I, 13) *AscI* and *Eag*I. The amount of DNA probes used to hybridse the YAC 761C2 PFGE filters was 50 ng for vector arm probes and for BAC DNA probes the amount of DNA required to visualise the hybridsed bands was standardised to 250 ngs. The YAC filters probed with vector probes were exposed to autoradiographs at -70^{o}C for overnight, where as YAC filters probed with BAC DNA probes were required to be kept at -70^{o}C for 4 to 7 overnights in order to visualise the hybridised bands. The restriction map constructed from these hybridization results is presented in Figure 4.4.

Conditions for PFGE were: gel: 1% agarose gel in 0.5X TBE, runnning temperature: 4^{o}C, resolution limits: 50 kb to 500 kb, run time: 28 hours.

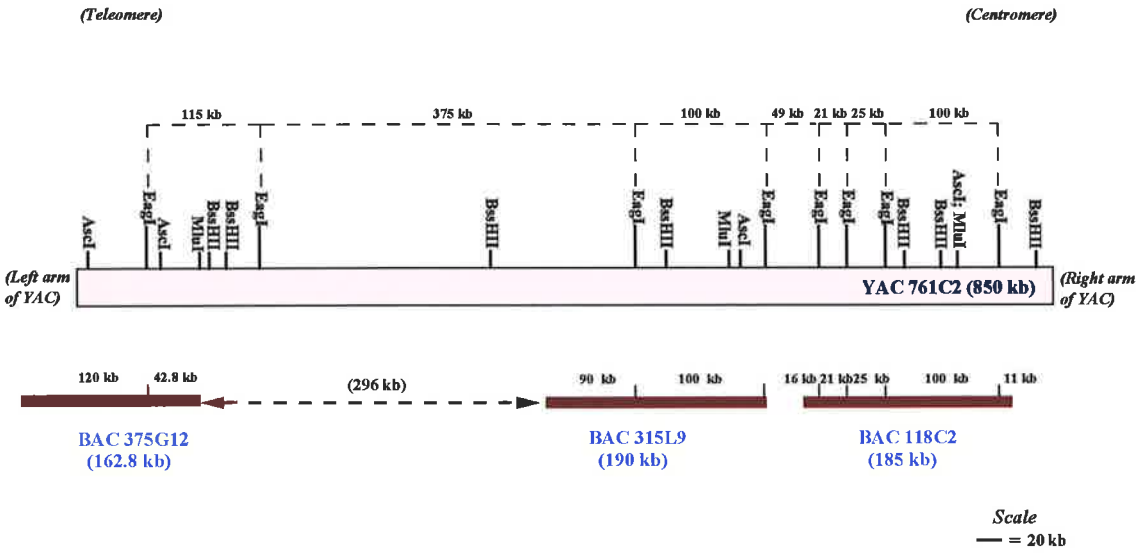


Figure 4.4 Restriction map of YAC 761C2. This map was prepared by using the hybridisation results obtained from probing YAC 761C2 PFGE filters with DNA probes: pUC19, pBR322, BAC 375G12, BAC 315L9 and BAC 118C2. The restriction fragments hybridised to pUC19 and pBR322 probes were located to the left and right ends of the YAC 761C2 respectively. The relative order of the restriction fragments detected within YAC 761C2 using the BAC probes is based on their arrangement with respect to the end fragments of YAC 761C2. The sizes for *Eag*I restriction fragments are indicated.

The Southern hybridisation results obtained from probing the YAC 761C2 filters with left and right vector arm probes pUC19 and pBR322 respectively and all three human BAC probes are summarised in Figure 4.4. The restriction fragments hybridised to pUC19 and pBR322 probes were located to the left and right end of YAC 761C2 respectively. The BACs 375G12, 315L9 and 118C2 helped to determine the restriction fragments to which they were contained. The relative order of the restriction fragments detected within the YAC by using BAC probes is based on their arrangement with respect to the ends fragments detected in YAC 761C2. The localisation of BACs to YAC restriction map was also confirmed by aligning the *Eag*I restriction fragments in BACs (Figure 4.2) to *Eag*I restriction sites detected within YAC 761C2. The sizes of *Eag*I restriction fragments in YAC 761C2 and BACs 375G12, 315L9 and 118C2 are indicated in Figure 4.4.

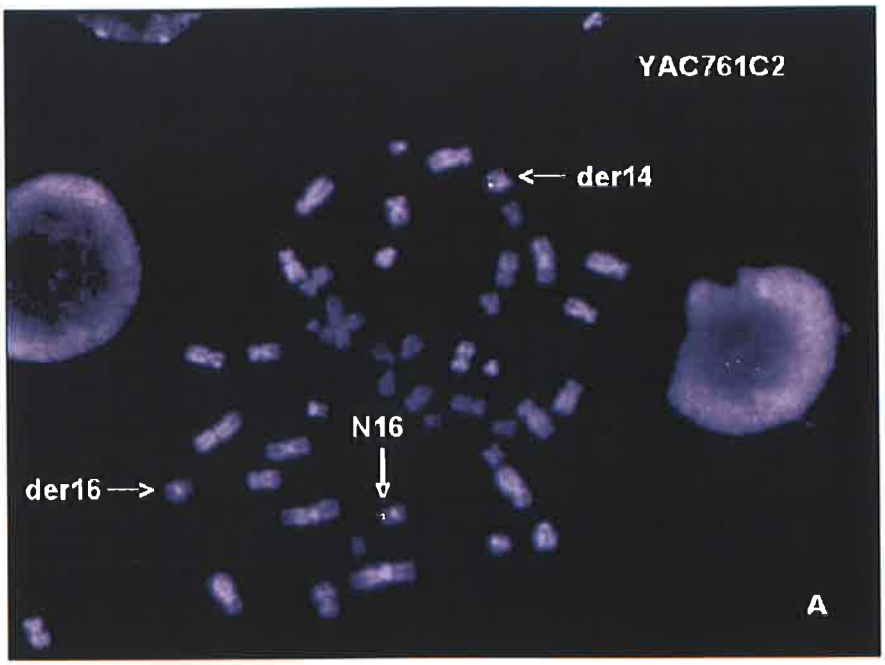
4.3.5 FISH analysis of DNA clones YAC 761C2, BACs 375G12, 315L9 and 118C2

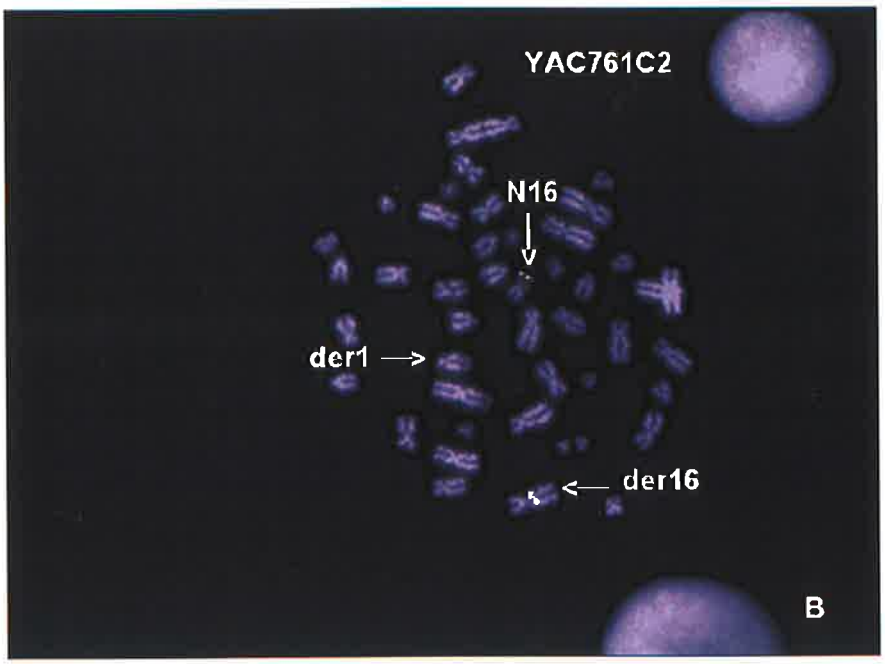
In order to determine the location of the 16p13.3 breakpoints of t(1;16) contained in CY196 and t(14;16) contained in CY182 with respect to the YAC restriction map (Figure 4.4), the DNA of clones YAC 761C2 and BACs 375G12, 315L9 and 118C2 were used for FISH analysis of metaphase chromosomes with the translocation (Figure 4.5). The data revealed that the YAC 761C2 was distal to the 16p13.3 breakpoint of the t(14;16) (Figure 4.5 A) and proximal to the 16p13.3 breakpoint of the t(1;16) (Figure 4.5 B). The BACs 315L9 and 118C2 also showed proximal signals to the 16p13.3 breakpoint of the t(1;16). An example of the proximal signal to the 16p13.3 breakpoint of the t(1;16) is shown in Figure 4.5 B. The BAC 375G12

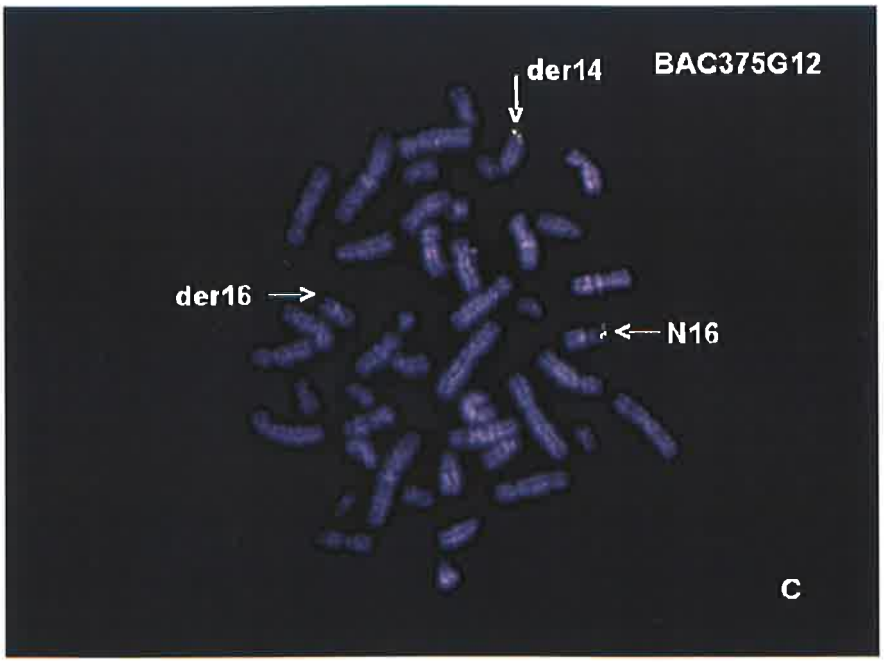
showed distal signals to both of the 16p13.3 breakpoints of the t(14;16) and the t(1;16) (Figure 4.5 C and D). The BACs 315L9 and 118C2 showed FISH signals on the normal chromosome 16 and the derivative 14 of the t(14;16) indicating that these clones were also distal to the 16p13.3 breakpoint of the t(14;16). An example of distal signal obtained from all three BAC DNA clones 375G12, 315L9 and 118C2 with the cell line of t(14;16) is presented in Figure 4.5 C. The FISH results for the DNA clones YAC 761C2 and BACs 375G12, 315L9 and 118C2 with respect to the 16p13.3 breakpoint of the t(1;16) and t(14;16) together with the restriction map data presented in Figure 4.4 are summarised in Figure 4.6. The results of FISH analysis together with YAC restriction map data helped in localisation of the two 16p13.3 breakpoints (Figure 4.6)

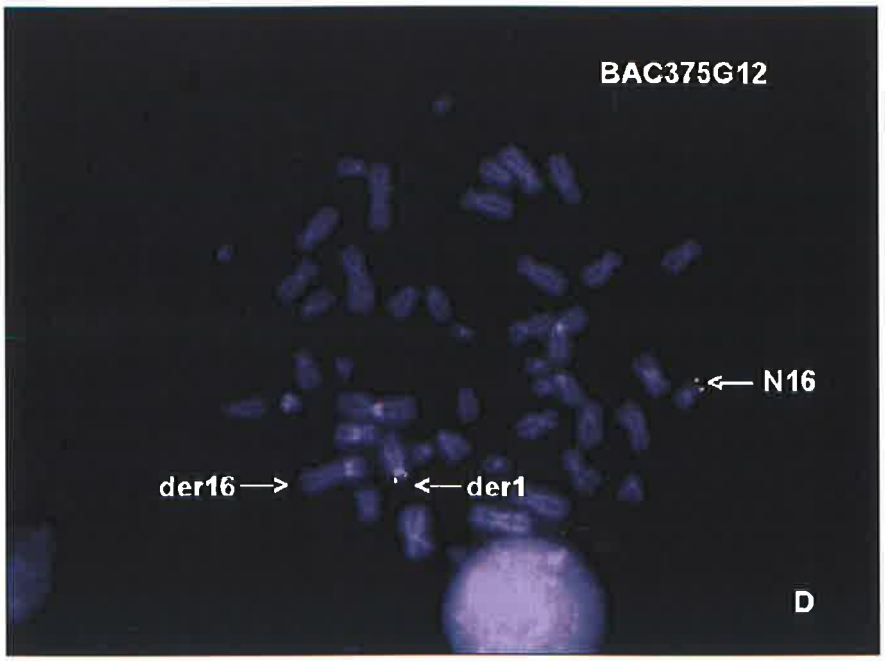
FISH analysis suggested that the 16p13.3 breakpoint of t(1;16) was proximal to the BAC 375G12 and distal to the BAC 315L9. These data when taken together with the YAC restriction map data indicate that this breakpoint lies in the 296 kb gap between the BACs 375G12 and 315L9 (Figure 4.6). Additional data generated with the cloning of 16p13.3 breakpoint of the t(1;16) (Chapter 5) further narrowed this breakpoint interval to 66 kb. However, the FISH results for YAC 761C2 (Figure 4.5 B) revealed that the YAC was proximal to this breakpoint. This proximal FISH signal was possibly obtained because the major length of the YAC 761C2 is proximal to the 16p13.3 breakpoint of the t(1;16) (Figure 4.6). The FISH analysis data together with the YAC restriction map data separated the two 16p13.3 breakpoints by a distance greater than at least 758 kb (Figure 4.6).

Figure 4.5 Fluorescent *in situ* hybridisation. **(A)** The hybridisation signals from the YAC 761C2 on the normal chromosome 16 and the derivative 14 of the t(14;16), indicating that YAC 761C2 is distal to the 16p13.3 breakpoint of t(14;16). **(B)** The hybridisation signals from the YAC 761C2 on the normal chromosome 16 and the derivative 16 of the t(1;16) indicating that YAC 761C2 is proximal to the 16p13.3 breakpoint of t(1;16). The BAC DNA clones 315L9 and 118C2 also gave proximal signal to the 16p13.3 breakpoint of t(1;16). **(C)** An example of the hybridisation signals on the normal chromosome 16 and the derivative 14 of the t(14;16) obtained from hybridising BAC DNA clones 375G12, 315L9 and 118C2 to the metaphase chromosomes of the patient with t(14;16). All three BACs were located distal to the 16p13.3 breakpoint of t(14;16). **(D)** The hybridisation signals from the BAC 375G12 on the normal chromosome 16 and the derivative 1 of the t(1;16) indicating that BAC 375G12 is distal to the 16p13.3 breakpoint of t(1;16).


YAC761C2
der14
N16
der16
A


YAC761C2
N16
der1
der16
B


der14
BAC375G12
der16
N16
C


BAC375G12
N16
der16
der1
D

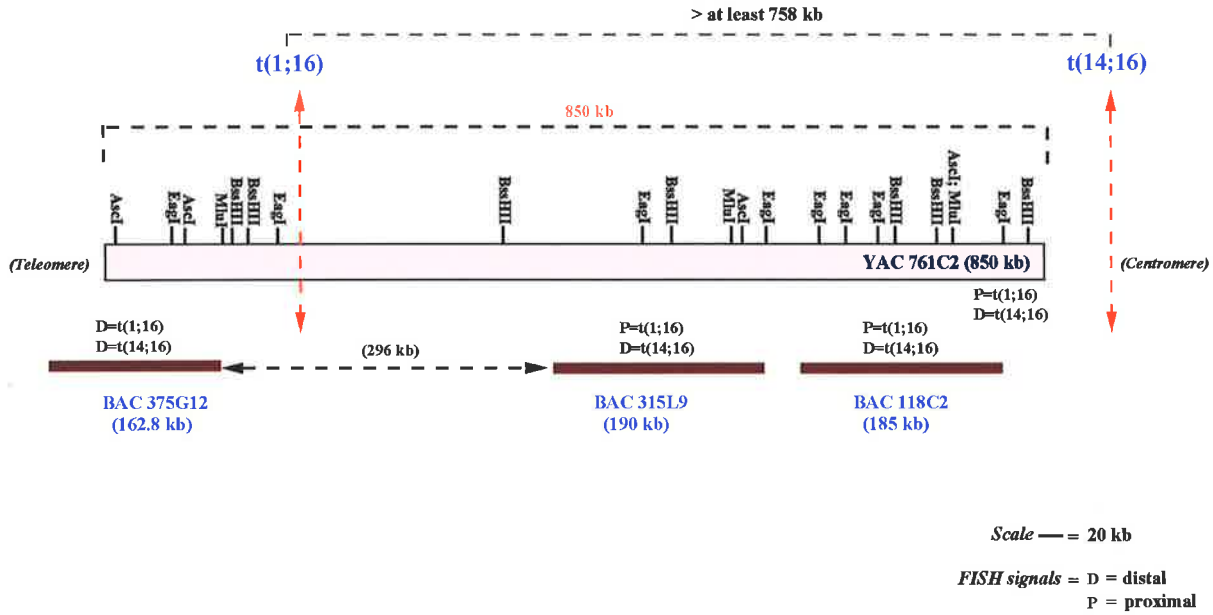


Figure 4.6 Localization of the 16p13.3 translocation breakpoints with respect to the YAC restriction map. The figure summarises the YAC 761C2 restriction map data and FISH results for DNA clones YAC 761C2 and BACs 375G12, 315L9 and 118C2 obtained from the hybridisation of these clones to the metaphase chromosomes of the two patients with t(1;16) and t(14;16). The FISH signals for respective clones with respect to the 16p13.3 breakpoints are indicated. The data when taken together separated the two 16p13.3 breakpoints by a distance greater than at least 758 kb and placed the 16p13.3 breakpoint of t(1;16) in the 296 kb gap between the BACs 375G12 and 315L9. Additional data generated from the cloning of 16p13.3 breakpoints of t(1;16) (presented in Chapter 5) and t(14;16) (presented in Chapter 6) revealed that the distance between the two breakpoints is at least a megabase.

4.3.6 Mapping of the C16orf5 gene with respect to the YAC 761C2 restriction map.

With the localisation of the two 16p13.3 breakpoints, it was possible to determine if C16*orf*5 gene was disrupted by the 16p13.3 breakpoint of the t(1;16). The SP6 and T7 end probes (Table 4.1) generated from the DNA clones spanning the C16*orf*5 gene were used for mapping experiments. The position of the C16*orf*5 gene with respect to the YAC map was estimated by PCR analysis of DNA clones YAC 761C2 and BACs 375G12, 315L9 and 118C2 used to construct the YAC restriction map of the breakpoints region.

The SP6 and T7 end probes were generated (section 2.3.7.4.3) from PAC 102J11. The PAC 102J11 was selected for generation of the end probes because FISH analysis with this clone on the metaphase chromosomes of the two patients with t(1;16) and t(14;16) (section, 3.3.8 9) (Figure 3.5 C and D) showed signal distal to both the 16p13.3 translocation breakpoints. However, when sequence was obtained for both ends of PAC 102J11 it was only possible to design the primers for the SP6 end as the sequence for the T7 end contained Alu repeat sequences. Therefore, two other overlapping PAC clones 43I06 and 795L6 (Figure 3.11), which also contained the entire gene, were then used to generate the SP6 and T7 end sequences of the inserts. Primers for PCR mapping were designed from the generated sequences. The sequences for primers designed for the SP6 end of PAC 102J11, and the SP6 and T7 ends of PACs 43I06 and 795L6, are presented in Table 4.1.

Table 4.1 Primers sequences for end probes generated from DNA clones spanning the C16*orf5* gene for PCR mapping.

Primer name	Primer sequence 5'-3'		PCR product size (bp)
	Forward primer	*Reverse primer*	
PAC 102J11 SP6	**ctg cct cag cct cct gag ta**	**gag ttc gag acc agc ctg ac**	115
PAC 43I06 SP6	**tgc att gac tgc aaa aga gg**	**agt cta ggc ctt ggg aga gc**	133
PAC 43I06 T7	**gca aca gag caa gta ctc c**	**act cct gac ctt gtg atc cg**	122
PAC 795L6 SP6	**ggc gac aga ctg gaa gaa aa**	**cac gaa agt tcc aga gag gg**	118
PAC 795L6 T7	**cgt gca cat atc caa cca ag**	**gcc tgg gag aca aaa tga aa**	181

The results of PCR mapping of SP6 and T7 end probes designed from the DNA clones spanning the C16*orf5* gene are shown in Table 4.2. The PCR mapping of the SP6 and T7 end probes with respect to the YAC restriction map placed the C16*orf5* gene distal to YAC 761C2 and 16p13.3 breakpoint contained in CY196, in the hybrid interval defined by 23HA and CY196 (Table 4.2 and Figure 4.7).

In an attempt to estimate the distance between the contig spanning the C16*orf5* gene and the 16p13.3 breakpoint of the t(1;16) (Figure 4.7), BLAST searches (May 17, 2000-November 8, 2000) were carried to link the gap between BAC 375G12 and the C16*orf5* contig (Figure 4.7). BLAST searches with the BAC 375G12 sequence (accession number, AC005751) through htgs (high throughout genome sequencing) database at NCBI identified the following overlapping BAC clones in the order: BAC 375G12-BAC 19H6-BAC 97H22-BAC124K4-BAC 61L4. However, the sequence for these clones [BACs 19H6, 97H22, 124K4 and 61L4, (RPCI-11 Human BAC library)] were only available as unordered pieces (contigs), and the relative

orientation of the contigs as not known. Furthermore, the end contigs in the sequence of BAC clones 124K4 and 61L4 were not placed at the extreme ends of the sequence. For example, when the SP6 (accession number, AQ342870) and T7 (accession number, AQ342873) end sequences of the BAC 124K4 were blasted against the 248 kb sequence of this BAC, the SP6 end was contained within the nucleotides 196,207-196,729 bp and T7 end sequence was contained within the nucleotides 196,836-196,927 bp and 197,140-197,423 bp. This indicated that the SP6 and T7 ends were present close to each other instead of being present at the extreme ends of the 248 kb sequence of the BAC 124K4. Also for the clone BAC 61L4 the SP6 end sequence (accession number, AQ202186) was present within the nucleotides 164,567-164,212 of the 255 kb sequence available, instead of being present in the extreme end. The SP6 and T7 end sequences for the clone BAC 19H6 and T7 end for the clone BAC 97H22 were not available in the database.

Although, BLAST searches involving the BAC 375G12 sequence identified the extending DNA clones, because the sequence for these clones as available in unordered contigs whose orientations were not known, it was difficult to estimate the distance extended between the BAC 375G12 and the most distal clone in the contig, BAC 61L4. Therefore, these clones are not incorporated in Figure 4.7. However, the sequence for BAC 375G12 (162 kb) was available as the finished sequence and it overlapped with BAC 19H6 sequence (208 kb) by 50 kb, suggesting that 158 kb of the sequence was extended by BAC 19H6 from the clone BAC 37512 (Figure 4.7). Also, the YAC map data taken together with the FISH analysis results (section 4.3.5) placed the 16p13.3 breakpoint of the t(1;16) in the 296 kb gap between the BACs

375G12 and 315L9 (Figure 4.6). The 16p13.3 breakpoint interval was further narrowed to 66 kb with the cloning of this breakpoint (Chapter 5). The BLAST search results of the BAC 37512 and 19H6 when taken together with the YAC map data, revealed that the distance of the C16*orf*5 gene from the 16p13.3 breakpoint of the t(1;16) is greater than at least 386 kb (Figure 4.7).

Further steps towards the extension of the constructed contig (using BAC 375G12 sequence) by screening the PAC/BAC libraries using BAC ends from the extending clone 61L4, to link the gap between BAC 375G12 and the C16*orf*5 contig, was not included in this project due to the time constraints. The orientation of the contig spanning the C16*orf*5 gene (Figure 4.7) may be reversed according to the orientation of the contained STSs in the radiation hybrid map.

4.3.7 Mapping of the DNA probes in the vicinity of the breakpoints at 16p13.3 to the YAC restriction map

The DNA markers, which were localised in the vicinity of the 16p13.3 breakpoints contained in the hybrids CY196 and CY182 were integrated to the YAC map by PCR analysis of the DNA clones YAC 761C2 and BACs 375G12, 315L9 and 118C2 to align the DNA markers with these clones. The results of PCR mapping are shown in Table 4.2 and are also summarised in Figure 4.7. Sequences of the primer pairs for the DNA markers listed in Table 4.2 are presented in Tables 3.1, 4.1 and 4.3.

4.3.7.1 Mapping of additional probes

Additional markers (D16S3042, 38H11, 25G3, SGC32612, D16S3135, D16S3128 and T53018) were identified to map within the hybrid interval defined by CY197 and CY182, thereby separating the chromosome breakpoints of these two hybrids. The primer sequences for these markers are presented in Table 4.3. These markers were localised to the short arm of chromosome 16 by PCR analysis of the somatic cell hybrids at 16p13.3 (Callen et al., 1995). An example of somatic cell hybrid analysis showing the localisation for a DNA marker (T53018) in the hybrid interval defined by CY197 and CY182 is presented in Figure 4.8. There were no markers identified earlier separating these two breakpoints (Figure 3.1). The order of the breakpoints of the hybrids (Figure 3.1) was refined from pter-CY196-CY197, CY177, CY182 to pter-CY196-CY197-CY182, CY177 (Figure 4.7). This placed the chromosome breakpoint contained in the hybrid CY197 between the 16p13.3 breakpoints of the two translocations contained in the hybrids CY196 and CY182. The CY197 hybrid was derived by fusing the human fibroblast cell line (K7006) containing the chromosome 16 translocation 46,XY,t(13;16)(q13.3;p13.3) with the A9 mouse cell line (Callen et al., 1990b). This was a familial translocation derived from cultured chronic villi cells.

Table 4.2 Mapping of the DNA probes in the vicinity of 16p13.3 breakpoints of t(1;16) and t(14;16) to the YAC restriction map.

DNA marker	Hybrid interval	YAC 761C2	BAC 375G12	BAC 315L9	BAC 118C2
SP6 PAC 102J11	23HA-CY196	-	-	-	-
SP6 PAC 43I06	23HA-CY196	-	-	-	-
T7 PAC 43I06	23HA-CY196	-	-	-	-
SP6 PAC 795L6	23HA-CY196	-	-	-	-
T7 PAC 795L6	23HA-CY196	-	-	-	-
D16S423	23HA-CY196	+	+	-	-
WI-4274	CY196-CY197	+	-	-	-
AFM339xg1	CY196-CY197	+	-	-	-
AFM288zf9	CY196-CY197	+	-	-	-
D16S3042	CY197-CY182	+	-	-	+
38H11	CY197-CY182	+	-	-	+
25G3	CY197-CY182	+	-	-	+
SGC32612	CY197-CY182	+	-	-	+
D16S3135	CY197-CY182	+	-	-	-
D16S3128	CY197-CY182	-	-	-	-
T53018	CY197-CY182	-	-	-	-

+ indicates the probe present in the respective DNA clones; -, probe is absent in the particular DNA clone. The DNA markers depicted in red separated the chromosome breakpoints in the two hybrids CY197 and CY182.

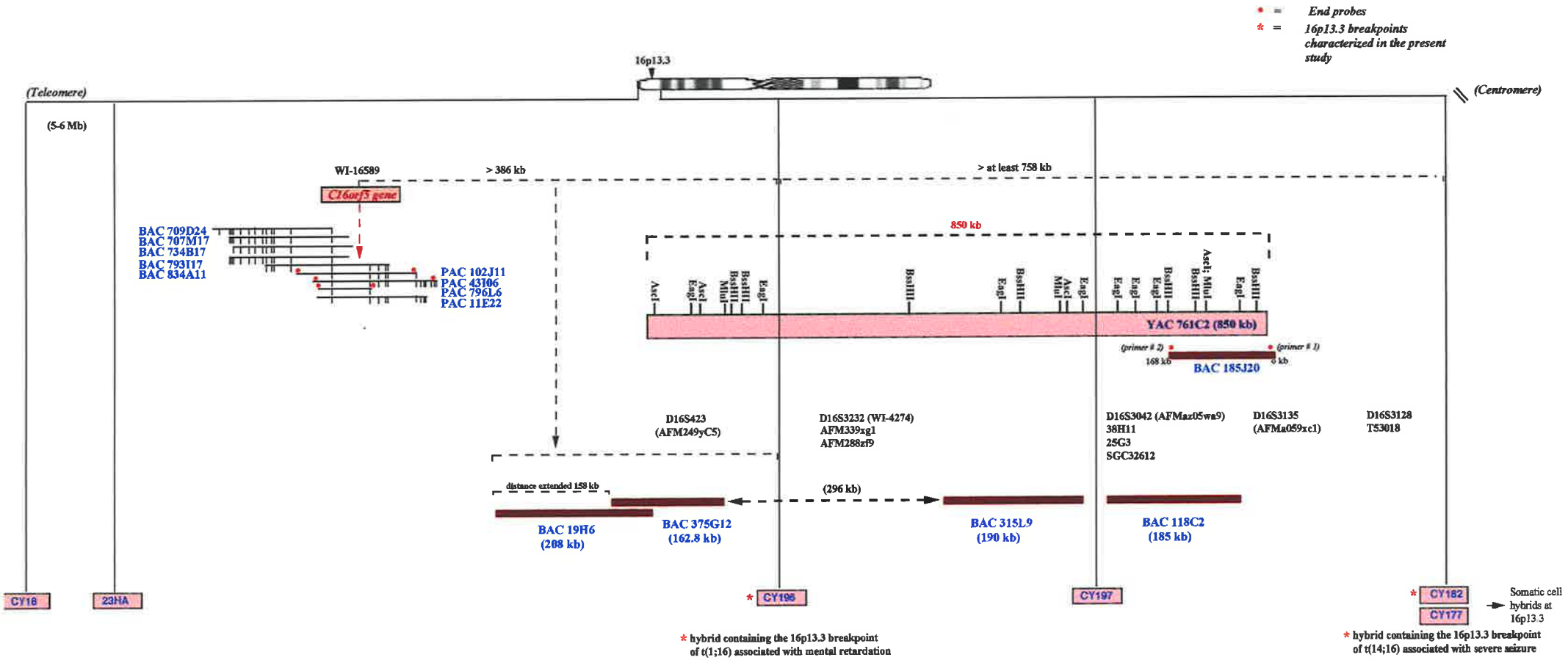


Figure 4.7 Long range integrated YAC map of the breakpoints region at 16p13.3. Integration of markers localized in the vicinity of 16p13.3 breakpoints contained in the hybrids CY196 and CY182 to the YAC restriction map of the region and localization of C16orf5 transcript with respect to the 16p13.3 breakpoint of t(1;16) and YAC 761C2 by PCR mappping the end probes generated from the DNA clones spanning the gene. The YAC restriction map data together with the BLAST search analysis using the sequence of BAC 375G12 placed the C16orf5 gene at a distance >386 kb from the 16p13.3 breakpoint contained in the hybrid CY196 in the hybrid interval 23HA-CY196. The orientation of the C16*orf5* contig seems to be more likely to be other way from the radiation hybrid mapping and GenBank data. The physical mapping data presented in this Figure refined the order of hybrid breakpoints at 16p13.3 from that presented in Figure 3.1 to pter-23HA-CY196-CY197-CY182,CY177.

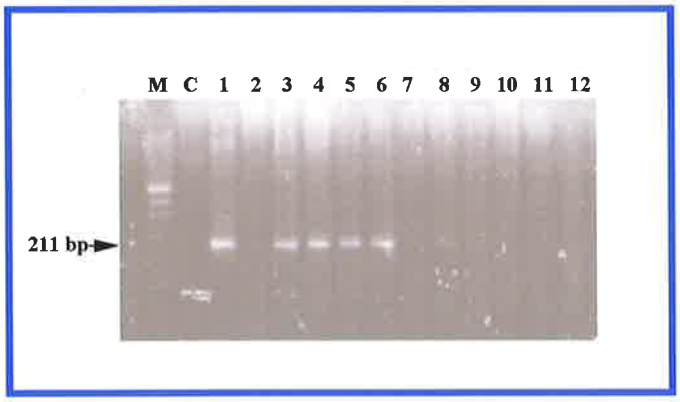


Figure 4.8 An example showing mapping of a DNA marker T53018 to the somatic hybrid interval CY197-CY182 at 16p13.3. Samples: M) pUC19 marker DNA, 1) Human DNA, 2) A9, 3) CY18, 4) 23HA, 5) CY196, 6) CY197, 7) CY182, 8) CY177, 9) CY194, 10) CY198, 11) CY168, 12) YAC 761C2. The CY18 and A9 somatic cell hybrids were used for determination of the chromosome 16 specific bands and mouse bands respectively. Lane C, represents no DNA (water) control. The DNA marker T53018 was located on chromosome 16 in the hybrid interval CY197-CY182 outside the YAC 761C2 (Figure 4.7).

Table 4.3 Primer sequences for the DNA markers physically mapped on the short arm of chromosome 16 within the hybrid interval defined by CY197 and CY182 at 16p13.3.

DNA markers	Primer sequences (5'-3')		PCR product size (base pairs)
	Forward primer	*Reverse primer*	
D16S3042	**agctttacgtggacaccaag**	**Ctacctatctgatcctagttgacc**	233
38H11	**tttgatgtaggtagccttttag**	**Gtcaacacatctgtcattcctt**	178
25G3	**aaggaatgtgtgccaagatgtg**	**Cagttcgctgttataggccatt**	103
SGC32612	**caaacagacctttggtttgag**	**Atgcagacaagcagagctgg**	150
D16S3135	**ctggaactggaactcatatc**	**Ccatagtgctgggacc**	178
D16S3128	**ggttgctgattccaggc**	**Cccgtgtatatgattatgtttgc**	174
T53018	**tggcagctgaaggctactct**	**Cctctcggatgggaactgta**	211

4.3.7.2 BAC 185J20

The DNA markers that were localised in the hybrid interval defined by CY197 and CY182 (Table 4.2) were used to identify the extending BAC/PAC clones in the htgs database to narrow the distance between the right arm of YAC 761C2 and the 16p13.3 breakpoint contained in CY182 (Figure 4.7). The htgs BLAST search at NCBI with DNA marker D16S3135 identified a BAC clone 185J20 (RPCI-11 Human BAC library). The sequence for this BAC was available in the database (accession number, AC007012), which enabled the design of primers (Table 4.4) from the extreme ends of BAC 185J20 to map this clone with respect to the YAC 761C2 and the 16p13.3 breakpoint contained in CY182. The BAC was localised on short arm of chromosome 16 in the hybrid interval defined by CY197 and CY182 (Figure 4.9 A) by PCR analysis of the somatic cell hybrids at 16p13.3.

Table 4.4 Primers designed from the extreme ends of BAC 158J20 sequence to map this clone with respect to YAC 761C2 and the 16p13.3 breakpoint contained in CY182.

Primer name	Primer sequence 5'-3'		PCR product size (bp)
185J20 #1			215
	Forward primer	**ccc agg cta cag gga caa ta**	
	Reverse primer	**aag agc cct tca gtc caa cc**	
185J20 #2			157
	Forward primer	**agc cat ttt gaa tgg tca ca**	
	Reverse primer	**aaa gag gac gct aga ccg ata**	

The BAC 158J20 ends were also mapped with respect to YAC 761C2. The primer pair (185J20 #1), which was derived form the first 200-600 bp sequence of BAC 185J20, was negative for YAC 761C2 (Figure 4.9 B). The primer pair (185J20 #2), which was present at 168 kb of the BAC 185J20 sequence, was positive for YAC 761C2 (Figure 4.9 B). The DNA marker D16S3135, which identified this BAC and was positive for YAC 761C2 (Table 4.2) and was located at 7kb of the BAC 185J20 sequence. PCR mapping of the end probes designed from BAC 185J20 sequence and DNA marker D16S3135 with respect to YAC 761C2 suggested that the BAC 185J20 extended the YAC map beyond the right arm of YAC 761C2 by approximately only 7 kb (Figure 4.7).

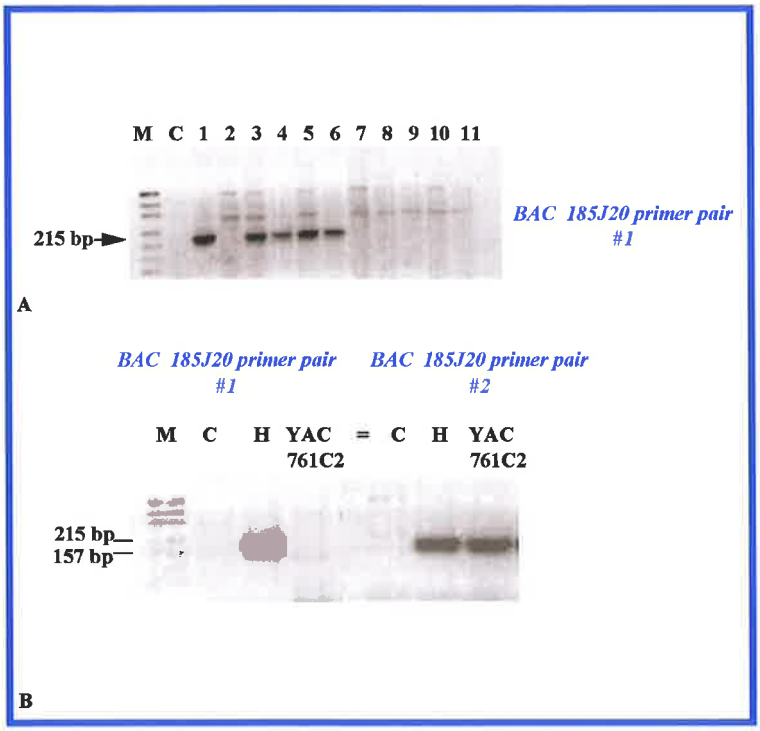


Figure 4.9 Localisation of the BAC 185J20. **(A)** An example showing localization of the primer pair designed from BAC 185J20 by PCR analysis of the somatic cell hybrids at 16p13.3. Samples: M) pUC19 marker DNA, 1) Human DNA, 2) A9, 3) CY18, 4) 23HA, 5) CY196, 6) CY197, 7) CY182, 8) CY177, 9) CY194, 10) CY198, 11) CY168. The CY18 and A9 somatic cell hybrids were used for determination of the chromosome 16 specific bands and mouse bands respectively. Lane C, represents no DNA (water) control. The BAC 185J20 was located on chromosome 16 in the hybrid interval CY197-CY182. **(B)** Localization of the primer pairs designed from the extreme ends of BAC 185J20 sequence with respect to YAC 761C2. M represents pUC19 marker DNA and = represents an empty lane. The sizes of the PCR products are indicated.

4.4 Discussion

The DNA markers D16S423 (AFM249yc5), AFM339xg1 and AFM288zf9 in the vicinity of the 16p13.3 breakpoints identified overlapping YAC clones YAC 761C2, YAC 929E9 and YAC 665G1, which were expected to span the hybrid interval defined by CY196 and CY182. These clones were obtained to prepare the long-range YAC map of the breakpoints region but only one of the clone YAC 761C2 was found to be of the correct size by PFGE to that given in the database of the Whitehead Institute. Therefore, a YAC restriction map was prepared using the YAC clone 761C2. The YAC DNA was digested and subsequently, YAC PFGE filters were prepared and probed with the left and right YAC vector arms probes, and human BAC DNA probes 375G12, 315L9 and 118C2 to determine the YAC restriction fragments.

FISH analysis of YAC 761C2 (850 kb) to the metaphase chromosomes of the two patients with t(1;16) and t(14;16) showed proximal signal to the 16p13.3 breakpoint of the t(1;16) and distal signal to the 16p13.3 breakpoint of the t(14;16). This suggested that the YAC 761C2 was located between the two 16p13.3 breakpoints of the t(1;16) and t(14;16). The FISH data for YAC 761C2 and BAC DNA clones 375G12, 315L9 and 118C2 together with the YAC restriction map separated the two 16p13.3 breakpoints contained in CY196 and CY182 by a distance greater than 758 kb. Later in the course of study, with the cloning of the 16p13.3 breakpoints of the t(1;16) and t(14;16), the distance between the two breakpoints was determined to be at least one megabase. The YAC map of the breakpoints region at 16p13.3 placed the

16p13.3 breakpoint of the t(1;16) in the 296 kb gap between the BACs 375G12 and 315L9 (Figure 4.6), which facilitated the positional cloning of this breakpoint.

The preliminary physical mapping data presented in Chapter 3, placed the C16*orf*5 gene distal to the 16p13.3 breakpoints t(1;16) and t(14;16). However, it was not established whether the gene is directly disrupted by any of the 16p13.3 breakpoints contained in CY196 and CY182. A Long-range YAC restriction map of the breakpoints region constructed using DNA clones YAC 761C2, BAC 375G12, BAC 315L9 and BAC 118C2 localized the 16p13.3 breakpoints contained in CY196 and CY182. It was therefore possible to map the end probes from the DNA clones spanning the C16*orf*5 gene with respect to the 16p13.3 breakpoint of the t(1;16) and the YAC map of the region. The mapping data revealed that C16*orf*5 gene lies distal to YAC 761C2, in the hybrid interval defined by 23HA and CY196, at a distance greater than 386 kb from the 16p13.3 breakpoint of the t(1;16). The C16*orf*5 gene was not directly disrupted by the *de novo* translocations in the two patients with mental retardation and severe seizures.

The DNA probes in the vicinity of the 16p13.3 breakpoints of t(1;16) and t(14;16) were also integrated into the YAC restriction map (Table 4.2) to align these markers with the DNA clones YAC 761C2 and BACs 375G12, 315L9 and 118C2. Additional markers (D16S3042, 38H11, 25G3, SGC32612, D16S3135, D16S3128 and T53018) were identified to map between the hybrid interval defined by CY197 and CY182, there by separating the breakpoints in these two hybrids. There were no markers identified earlier separating these two breakpoints in the previously reported map of

chromosome 16 (Callen et al., 1995), and in the data presented and discussed in Chapter 3, Figure 3.1. DNA markers localised within the hybrid interval defined by CY197 and CY182 were used to identify the extending BAC/PAC clones in the database to narrow the distance between the right arm of YAC 761C2 and the 16p13.3 breakpoint contained in CY182. The DNA marker D16S3135 identified a BAC clone 185J20 that was localised on short arm of chromosome 16 in the hybrid interval defined by CY197 and CY182 by PCR analysis of the somatic cell hybrids at 16p13.3. However, PCR mapping of the end probes designed from BAC 185J20 sequence with respect to YAC 761C2 suggested that this BAC clone extended the YAC map at the right arm of YAC 761C2 by approximately only 7 kb.

To further refine the two-translocation breakpoints contained in CY196 and CY182, in order to identify if any transcript is directly disrupted at the 16p13.3 breakpoints of t(1;16) and t(14;16), a BAC/PAC contig was constructed across both the breakpoints. The cloning of 16p13.3 breakpoints of the t(1;16) and t(14;16) is presented in Chapters 5 and 6 respectively.

Positional Cloning of the t(1;16) associated with Severe Mental Retardation

Chapter 5

5.1 Introduction

Mental retardation (MR) is a very common phenotype and is known to be caused by defects in a large number of genes (Antonarakis and Aelst, 1998). Approximately 2%-3% of the human population has either learning or behaviour disabilities with an intelligence quotient (IQ) less than 70 and 0.3% of the individuals are severely handicapped (IQ<50) (McLaren and Bryson, 1987). The disorder has a substantial genetic component and there may be a genetic cause in approximately 50% of severely retarded patients (Hagberg and Kyllerman, 1983).

Identification of families with mentally retarded males has led to the mapping and cloning of a number of X-linked genes for mental retardation (section 1.8.1.1). However, there has been little progress in the identification of equivalent genes on autosomes (section, 1.8.1.2). Although, the positional cloning efforts based on the investigation of autosomal translocations, deletion mapping and candidate gene strategy have identified many genes involved in the syndromal forms of autosomal MR (section 1.8.1.2). However, these genes may not be directly involved in human intellectual and development of learning abilities as MR is a secondary feature in most of the syndromes (Chelly, 1999). Further, there is no gene identified yet for autosomal non-specific MR. Therefore, it is important to exploit any opportunity to clone autosomal genes for non-specific MR. The identification of autosomal genes implicated in MR can be accelerated by thorough investigation of molecular abnormalities involving autosomes that are associated with MR.

Balanced structural rearrangements of the autosomes have been observed in 0.15% of the live-born human infants and are usually associated with a normal

clinical phenotype (Tharapel et al., 1977). However, the risk of serious congenital anomaly is estimated to be 6.1% (n=163) for *de novo* reciprocal translocations (Warburton, 1991), which is considered to arise from such translocation disrupting a gene at or near the translocation breakpoint (Bedell et al., 1996; Baere et al., 2000). A patient was reported at the Department of Cytogenetics, WCH, Adelaide, with a balanced *de novo* translocation 46,XY,t(1;16)(q12;p13.3) associated with severe mental retardation. This provided a unique resource for the application of the positional cloning strategy to identify an autosomal gene for non-specific mental retardation.

The search for the gene responsible for the patient's phenotype was focussed on chromosome 16 because the chromosome 1 breakpoint was shown to be within the 1q heterochromatin. As described in the previous chapter the 16p13.3 breakpoint of the t(1;16) was narrowed to a 296 kb distance between the BACs 375G12 and 315L9 by use of a YAC restriction map of this region.

The major aim of this Chapter involved the detailed molecular characterisation of the t(1;16)(q12;p13.3) *de novo* balanced translocation associated with severe mental retardation in order to identify if there was a gene disrupted by the breakpoint that was responsible for the patient's phenotype. In order to achieve this aim the breakpoint at 16p13.3 was cloned by constructing a contig of BAC/PAC clones. FISH analysis with these clones to metaphase chromosomes of the patient with the t(1;16) identified a BAC clone spanning the 16p13.3 breakpoint. To further narrow the region of 16p13.3 breakpoint to a smaller DNA clone a PAC contig was constructed across the 16p13.3 breakpoint. Subsequently, DNA of a PAC clone, which was found to span the 16p13.3 breakpoint was

sequenced. The breakpoint was localised within the PAC clone by PCR using the somatic cell hybrid CY196, which contains the der(16) of the t(1;16). The sequence data generated was analysed to identify any transcripts in the region.

5.2 Methods

Many of the methods used in this chapter are described in detail in Chapter 2. Specifically presented are the methods for the heterochromatin staining of the patient's metaphase chromosomes and the sequencing of the PAC 978I1 DNA clone.

5.2.1 Heterochromatin staining

Heterochromatin staining was performed by Helen Eyre (Department of Cytogenetics and Molecular Genetics, WCH, Adelaide). Metaphase chromosomes from the lymphoblastoid culture of the t(1;16) were stained by the Distamycin A/ DAPI method (Schweizer, 1981), which specifically stains the heterochromatic regions of chromosome 1, 9, 15, 16 and Y.

5.2.2 PAC/BAC DNA clones

The PAC clones were identified by hybridisation screening of the human PAC library RPC15 (PAC/BAC Resources, Buffalo, New York) with the 375G12 BAC end PCR product. The PAC library was screened as per the methods described in section 2.3.7.1. The positive PAC DNA clones were purchased from PAC/BAC Resources. These clones were confirmed as true positives by colony blot hybridisation (2.3.4.7.3). BAC 578P21 was obtained through Research Genetics (USA). DNA clones were grown at 37°C in LB (Luria Bertoni) medium (section 2.2.6.1) containing 50 ug/ml of kanamycin in the case of PACs or 30 ug/ml of

chloramphenicol for DNA cultures of the BACs. DNA was isolated using Qiagen 100 columns (section 2.3.1.3). The clone inserts were sized and restriction mapped using pulsed-field gel electrophoresis (PFGE) (section 2.3.4.4.3) and the infrequently cutting restriction enzymes *Not*I and *Eag*I (section 2.3.4.1). The gels were then transferred to Hybond-N^+ nylon membranes as described in section 2.3.4.7.1. The radiolabelling of probe, membrane hybridisation and washing was done as per the methods described in sections 2.3.4.

5.2.3 Somatic cell hybrid analysis

The somatic cell hybrid analysis was performed as described earlier in section 2.3.6. The primer pairs used in this Chapter were designed as described in section 2.3.5.1.

5.2.4 FISH analysis

FISH analysis of BAC 578P21 and PACs 978I1 and 1089J21 on the metaphase chromosomes of the patient with t(1;16) was performed by Helen Eyre (Department of Cytogenetics and Molecular Genetics, WCH, Adelaide) as described in section 2.3.11.

5.2.5 PAC 978I1 sequencing

The PAC 978I1, which was identified to span the 16p13.3 breakpoint of the t(1;16) was sequenced by the joint efforts of the Australian Genome Research Facility (Brisbane, Queensland), and Dr Gabriel Kremmidiotis and Alison Gardner (Department of Cytogenetics and Molecular Genetics, WCH, Adelaide). The DNA was prepared from PAC 978I1 and sheared by nebulization. Electrophoresis of the sheared DNA on agarose gel allowed isolation of DNA in

the 2-4 kb range. These fragments were cloned into the plasmid vector pUC 18 and used to transform competent *E. coli* cells. This work was performed by Dr Gabriel Kremmidiotis and subsequent colony isolation, DNA preparation and sequencing on an AB1377 sequencer were performed by the Australian Genome Research Facility (Brisbane, Queensland). Assembly and analysis of genomic sequence was performed together by Alison Gardner and Dr Gabriel Kremmidiotis using PHRED, PHRAP and GAP4 software (Ewing and Green, 1998) on a SUN workstation.

5.2.6 Computational analysis of PAC 978I1 sequence

The genomic sequence was masked for repeats (Repeat Masker, Baylor College of Medicine, http://dot.imgen.bcm.tmc.edu:9331) and blasted through dbEST at NCBI (http://ncbi.nlm.nih.gov) for the prediction of ESTs using the BLASTN algorithm (Altschul, 1990). The computer prediction of ESTs and exons was also done using the NIX program (UK HGMP Resource Centre, http://www.hgmp.mrc.ac.uk/). The PAC 978I1 sequence was also analysed for matches to the PROSITE database (http://www.expasy.ch/cgi-bin/scanprosite?1) and the ProDom NCBI-BLASTP2 database (http://protein.toulouse.inra.fr/prodom/cgi-bin/NewBlastProdomll.pl) for the identification of any motifs or domains in the sequence.

5.2.7 DNA sequencing of PAC ends, PCR, and RT-PCR products

PAC/BAC end sequences were generated by direct sequencing of the inserts using standard SP6 and T7 sequencing primers (section 2.3.7.4.2). Other PCR/RT-PCR products were either directly sequenced after DNA purification (section 2.3.2) using the BigDye sequencing chemistry (section 2.3.12.3) or subcloned into

pGEM-T vector (section 2.3.3) and sequenced using Dye Primer sequencing kit (section 2.3.12.1).

5.2.8 RT-PCR and Multiple Tissue Northern (MTN^{TM})analyses

The RT-PCR and Northern analyses were done as described in section 2.3.8 and 2.3.9 respectively to determine if ESTs and computer predicted exons on the PAC 978I1 sequence corresponded to expressed transcripts. A human RNA Master blot (Clontech) was probed with the insert of AI243740 as described in section 2.3.10. The RNA samples for RT-PCR analyses performed in section 5.3.4.2 were kindly provided by Dr Scott Whitmore (Department of Cytogenetics and Molecular Genetics, WCH, Adelaide).

5.3 Results

5.3.1 Case report of the patient with t(1;16)

Phenotype of the patient:

The patient presented at the age of four years and eleven months. He was born at 36 weeks after an uneventful pregnancy and was reported to have a single episode of fitting at 5 days of age. This boy had severe intellectual retardation with developmental skills at the 2-2$^{1/2}$ year level. Strabismus and down-turned angles of the mouth were the only dysmorphism. The family history revealed that the mother of the patient had probable intellectual retardation and father had learning difficulties up to Grade 6 (in the Netherlands). He was referred to the Victorian Clinical Genetics Service at 5 years of age where developmental delay and no clear dysmorphism were noted.

Karyotype of the patient:

The karyotype of the patient was 46,XY,t(1;16)(q12;p13.3). This translocation was *de novo* as parental karyotypes were normal. A somatic cell hybrid designated as CY196 (Callen et al., 1990b) was generated containing the derivative chromosome 16 of this translocation.

5.3.2 1q12 breakpoint

Distamycin A/DAPI (Sigma) staining of this patient showed that the derivative chromosome 16 had two positive regions (Figure 5.1 A). One was the pericentromeric heterochromatin of 16 and the other was a segment of the chromosome 1 pericentric heterochromatin translocated to 16p13.3. This demonstrates that the breakpoint of the chromosome one had occurred within the pericentric heterochromatin at 1q12.

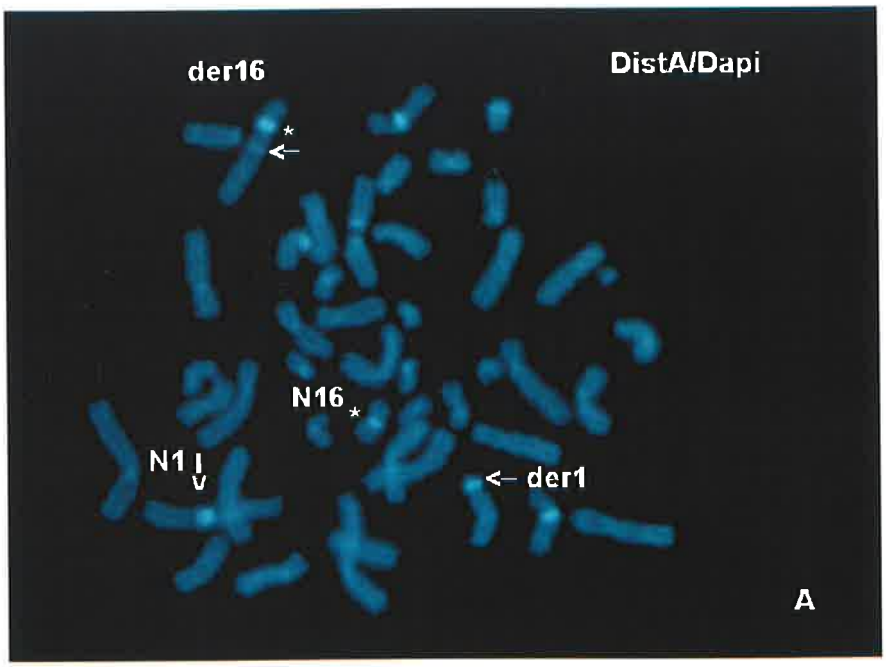
5.3.3 Cloning of the 16p13.3 breakpoint

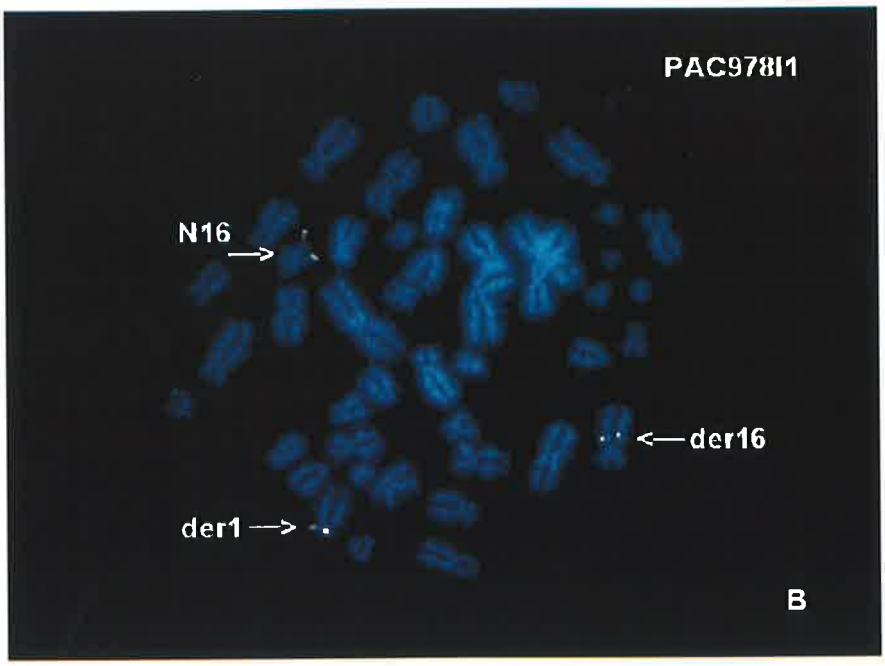
Since the chromosome 1 breakpoint of the t(1;16) was shown to be within the 1q heterochromatin and this region is considered to be devoid of genes, the approach was to clone the other breakpoint of the translocation at 16p13.3 in order to identify if this breakpoint disrupts a gene which may be responsible for the patient's phenotype.

5.3.3.1 BAC 375G12

The long range YAC map data presented in Figure 4.7 narrowed the 16p13.3 breakpoint of the t(1;16) to the 296 kb distance between the BACs 375G12 and 315L9.

Figure 5.1 Fluorescent *in situ* hybridisation. **(A)** Heterochromatin staining: distamycin A/DAPI staining showing the chromosome 1 breakpoint is within the heterochromatin. Chromosome 1 is indicated with arrows and chromosome 16 heterochromatin is indicated with asterisk. **(B)** An example of the spanning FISH signal to the 16p13.3 breakpoint of t(1;16). The FISH signals obtained by the hybridisation of PAC 978I1 to the metaphase chromosomes of patient were seen on the normal chromosome 16, the derivative 16 and the derivative 1 of the t(1;16) indicating that the PAC 978I1 spans the 16p13.3 breakpoint of t(1;16).

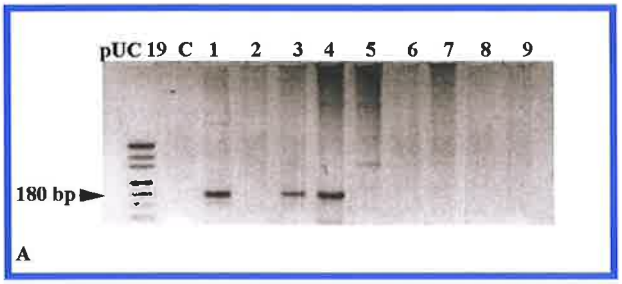

der16
DistA/Dapi
N16
N1
der1
A


PAC978I1
N16
der16
der1
B

To further reduce the breakpoint interval, the sequence of BAC 375G12, which was available in the database under the accession number AC005751, was used to generate end probes in order to identify overlapping DNA clones. These clones would likely be positioned in the 296 kb gap. The orientation of the BAC 375G12 with respect to the YAC 761C2 was determined by mapping its end probes. The end of BAC 375G12 contained within the YAC 761C2 is closest to the translocation breakpoint and was used to screen a PAC library.

Primers were designed from the extreme ends of the BAC 375G12 sequence, and are listed in Table 5.1. Information regarding SP6 and T7 ends sequences of this BAC was not available in GenBank. Therefore, the primers designed from 25 bp-317 bp of the 162.8 kb BAC sequence were designated as the "top end" primer pair and the primers designed from 162,025 bp-162,204 bp of the 162.8 kb BAC sequence were designated as the "bottom end" primer pair. The "top end" and "bottom end" primer pairs will be referred in the text as distal and proximal primer pairs respectively. It was not possible to design primers at the extreme proximal end of the sequence as the region between 162,360 bp- 162,820 bp contained repeat sequence.

Firstly, the localisation of these primer pairs on chromosome 16 was confirmed by PCR analysis of somatic cell hybrids with breakpoints at 16p13.3 (Callen et al., 1995). Both the primer pairs amplified the chromosome 16 specific band from the somatic cell hybrid CY18, which contained the entire chromosome 16 as the only human chromosome, and were localised distal to the 16p13.3 breakpoint contained in CY196.



BAC 375G12 proximal end primer pair

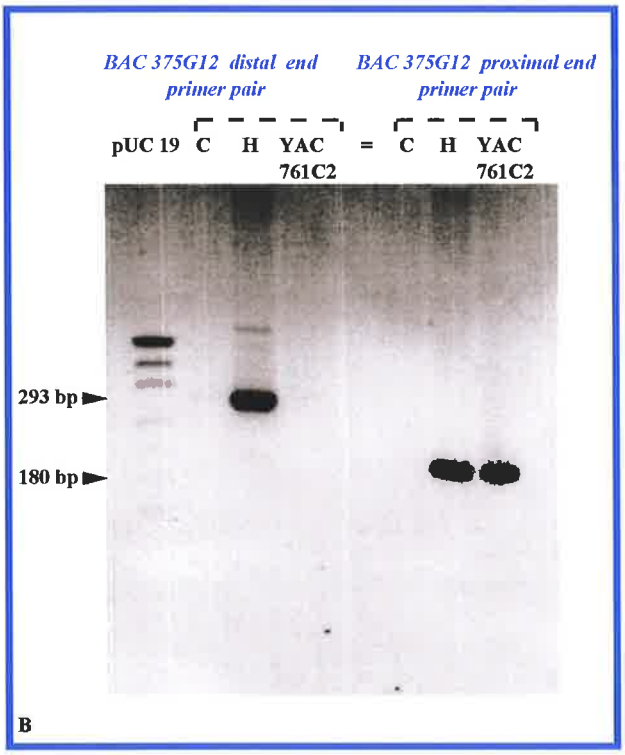


Figure 5.2 Mapping of distal and proximal ends of BAC 375G12. **(A)** An example of mapping of BAC 375G12 proximal end primer pair to chromosome 16 with respect to the 16p13.3 breakpoint in CY196 using somatic cell hybrids at 16p13.3. Samples: 1) Human DNA, 2) A9, 3) CY18, 4) 23HA, 5) CY196, 6) CY197, 7) CY182, 8) CY198, 9) CY168. The CY18 and A9 somatic cell hybrids were used for determination of the chromosome 16 specific bands and mouse bands respectively. Lane C, represents no DNA (water) control. The BAC 375G12 primer pairs were located on chromosome 16 in the hybrid interval defined by 23HA and CY196. **(B)** Mapping of distal and proximal ends primer pairs designed from BAC 375G12 with respect to YAC 761C2. The YAC 761C2 was found to contian the bottom end of BAC 375G12. Lane C, represents no DNA (water) control and H, positive control (human DNA from blood bank). The PCR product sizes are indicated by arrows. =: represents

An example showing localisation of these primer pairs to chromosome 16 in the hybrid interval defined by 23HA and CY196 is shown in Figure 5.2 A. PCR analysis of the YAC 761C2, with these primer pairs (Table 5.1) showed that the proximal end primers of the BAC 375G12 sequence were contained in the YAC 761C2 (Figure 5.2 B). The proximal end primers were therefore closest to the translocation breakpoint and were then used to screen a PAC library (see 5.3.3.6). While this was in progress further in silico analysis of the BAC 375G12 sequence was undertaken.

Table 5.1 Primer pairs designed from BAC 375G12 sequence (accession number, AC005751).

Primer name	Primer sequence (5'-3')	PCR product size (bp)	Positions of primers in BAC sequence
375G12/top end		293	25-317 bp
Forward primer	**GAC GGC CAT TCT CTG TGT CT**		
Reverse primer	**CGG GAA GGC TTT AAC TGT GA**		
375G12/bot end			
Forward primer	**TAT GAA AAA TGA CGC GTG GA**	180	162,025-162,204 bp
Reverse primer	**GCC ACT CCT TCC CCT AAG AC**		

'top' and 'bot' refer to the primers designed from the distal (top) and proximal (bottom) ends of the BAC 375G12 sequence.

5.3.3.2 Identification of BAC 578P21

The gss database (genome survey sequence database, which contains the sequences of BAC ends) was searched using the repeat masked BAC 375G12 sequence. The T7 end (AQ332080) of BAC 578P21 (RPCI11 Human Male BAC

Library) was identified. The BAC 578P21 was then obtained through Research Genetics and the DNA was prepared. The overlap and orientation of BAC 578P21 with respect to BAC 375G12 was determined by PCR analysis using the end primers (Table 5.1) of the BAC 375G12 sequence. The PCR results are shown in Figure 5.3. The BAC 578P21 overlapped only with the proximal end of the BAC 375G12, which indicated that this clone extended proximally from BAC 375G12 thereby potentially reducing the 296 kb gap at the 16p13.3 breakpoint interval. However, to estimate the distance extended by this clone from BAC 375G12 it was necessary to determine the size of this BAC by restriction fragment analysis followed by the identification of their common restriction fragments.

5.3.3.3 Estimation of the size of BAC 578P21 by PFGE.

The BAC 578P21 DNA was digested using the restriction enzymes *Not*I and *Eag*I and the DNA fragments resolved using PFGE (2.3.4.4.3) at resolution limits ranging from 5 kb-150 kb (Figure 5.4 A). The higher molecular weight restriction fragments were further resolved at resolution limits ranging from 130 kb-40 kb (Figure 5.4 B). By summation of these restriction fragments the size of the BAC 578P21 was estimated to be 147 kb by PFGE analysis.

5.3.3.4 Identification of the overlapping restriction fragments in BAC 375G12 and 578P21.

The PFGE filter-containing DNA restricted with enzymes *Not*I and *Eag*I was generated and probed with the α^{32}P-dCTP labelled BAC 375G12 proximal end probe. The results of the Southern hybridisation of this filter are shown in Figure 5.5 A. The α^{32}P-dCTP labelled BAC 375G12 end probe hybridised to itself and to a 5 kb restriction fragment of the BAC 578P21.

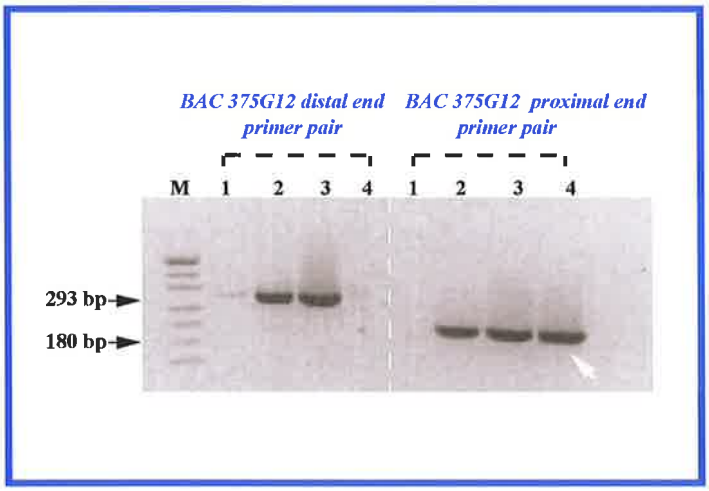


Figure 5.3 PCR analysis using BAC 375G12 distal and proximal end primer pairs to confirm the overlap between BAC 375G12 and BAC 578P21. Samples: 1) negative control (no DNA), 2) postive control (genomic DNA from blood bank), 3) BAC 375G12, 4) BAC 578P21. The BAC 375G12 proximal end primer pair was also positive for BAC 578P21, which is depicted by an arrow. This experiment indicated the overlap between BAC 375G12 and BAC 578P21 that was also confirmed by Southern analysis (refer Figure 5.5).

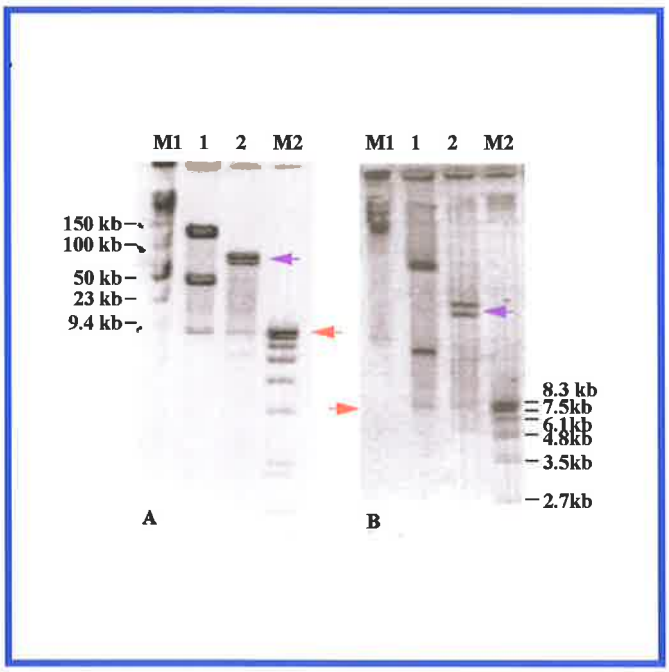


Figure 5.4 Estimation of the size of BAC 578P21 by PFGE. Samples: M1) low pulsed-field gel molecular weight marker 1) BAC 375G12, 2) BAC 578P21, M2) SPP1 marker. The sizes and positions of the marker bands are indicated. BAC 375G12 was used as a positive control in the experiment. DNA for BACs 375G12 and 578P21 was digested using restriction enzymes *Not*I and *Eag*I. Subsequently, the restriction digested DNA fragments were resolved by PFGE. The conditions for resolution of DNA fragments were: gel: 1% agarose gel in 0.5X TBE, running temperature: 14^{o}C, angles of electrodes: 120^{o}.

(A) The resolution of DNA fragments within the range 5 kb - 150 kb, time:16 hours. The purple colour arrow shows two DNA fragments for BAC 578P21, which were further resolved in gel B.

(B) The resolution of DNA fragments within the range 40 kb - 130 kb. The restriction fragments obtained for respective clones were added leaving the vector bands to estimate the sizes of the respective BAC DNA clones. The red arrows depicts the bands of the BAC vector pBACe3.6. The sizes of BACs as estimated by PFGE were: BAC 375G12, 162 kb and BAC 578P21, 146.7 kb.

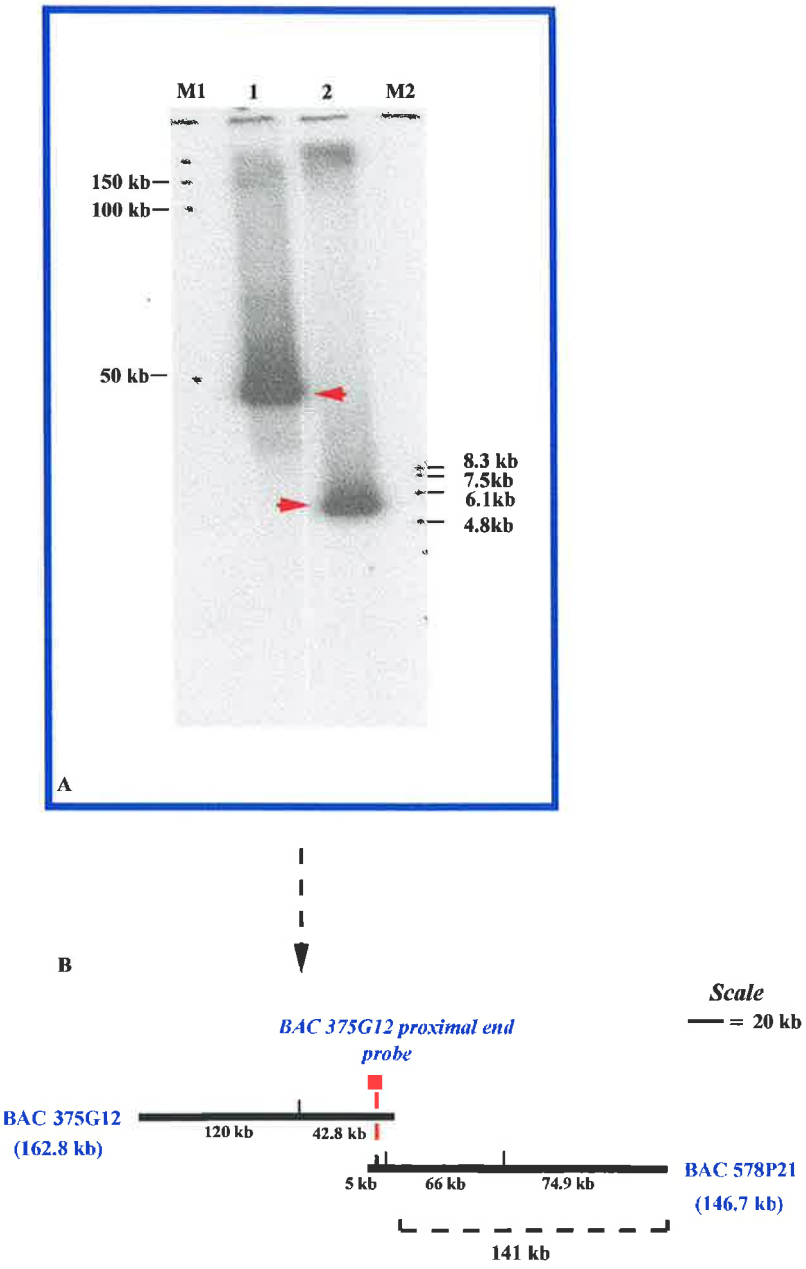


Figure 5.5 Southern hybridisation analysis to confirm the overlap between BACs 375G12 and 578P21. **(A)** Samples: M1) low pulsed-field gel molecular weight marker 1) BAC 375G12, 2) BAC 578P21, M2) SPP1 marker. The sizes and positions of the marker bands are indicated. BAC 375G12 was used as a positive control in the experiment. BACs 375G12 and 578P21 DNA was digested using restriction enzymes *Not*I and *Eag*I. The restriction enzyme digested DNA fragments were resolved by PFGE and subsequently, PFGE filter was made. The filter was probed with PCR product generated using BAC 375G12 proximal end primer pair. Autoradiography was carried out at room temperature for 3 to 4 hours to visualise the hybridised bands, which are depicted by red arrows. **(B)** This Figure summarises the results of Southern hybridisation shown in Figure 5.5 A. The position of end probe generated from BAC 375G12 is indicated. The DNA fragments positive for the probe are shown by a dashed red line. BAC 578P21 extended the distance from BAC 375G12 by 141 kb.

Therefore, the BAC 578P21 extended the length of BAC 375G12 by another 141 kb (Figure 5.5 B) and further narrowed the 296 kb gap between the BACs 375G12 and 315L9.

5.3.3.5 FISH analysis of BAC 578P21

The BAC 578P21 narrowed the gap between the BACs 375G12 and 315L9. This gap contains the 16p13.3 breakpoint of the t(1;16). Therefore, FISH analysis was carried out with this clone on metaphase chromosomes of the patient with t(1;16). FISH signals were obtained on the normal chromosome 16, and both the derivative 1 and the derivative 16 of the t(1;16). This indicated that this clone spans the 16p13.3 breakpoint. An example of the DNA clone (PAC 978I1) depicting spanning FISH signal on the metaphase chromosomes of t(1;16) is shown in the Figure 5.1 B.

5.3.3.6 Screening of the human PAC library (RPCI5)

Although the BAC clone 578P21 spanned the 16p13.3 breakpoint of the t(1;16), the sequence of this clone was not available in the GenBank at that time and the insert size as estimated by PFGE was 147 kb. Due to budgetary restraints it was necessary to further refine the breakpoint region to a smaller DNA clone, which could be more economically sequenced. Accordingly, a human PAC library (RPCI5) (PAC/BAC Resources, Buffalo, New York) was screened using the BAC 375G12 proximal end probe.

Seven positive PAC DNA clones were identified, Figure 5.6. These clones were purchased and confirmed as true positives by colony blot hybridisation using the same probe used to screen the PAC library.

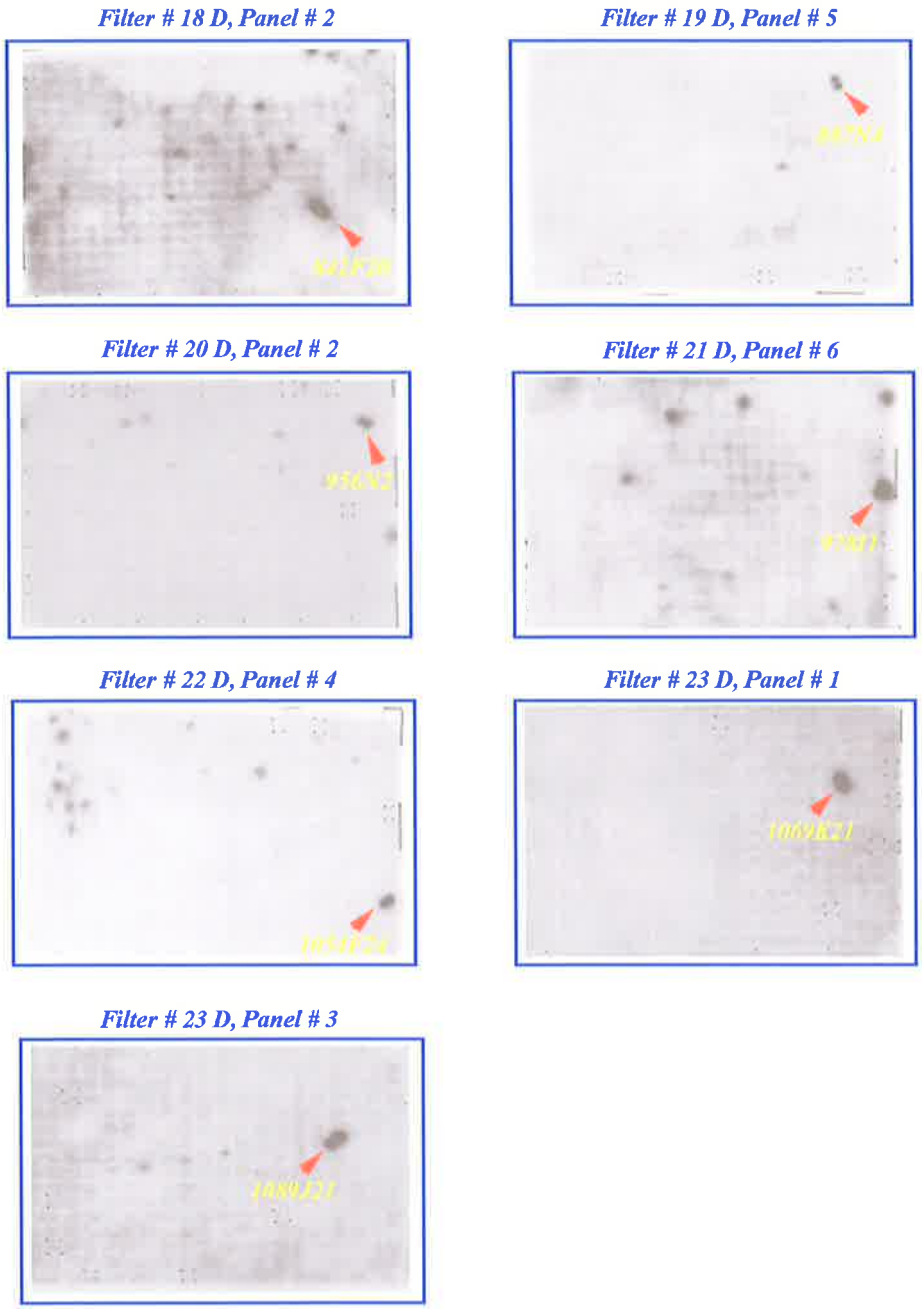


Figure 5.6 Screening of the human PAC library (RPCI5) with BAC 375G12 proximal end probe. The positive clones, which are indicated by red arrows were obtained through PAC/BAC Resources, Buffalo, New York. The double dots are positives as duplicate dots indicated by red arrows. Only the panels (field sections) for respective filters containing the positive clones are shown. The filter numbers and panel numbers for respective filters along with the names of positive clones identified in each filter are indicated.

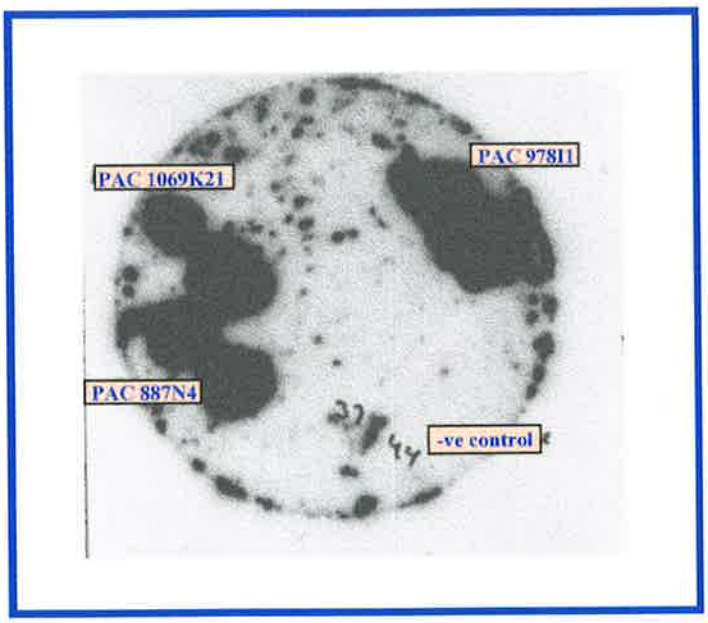


Figure 5.7 Colony blot hybridisation. The PAC DNA clones (Figure 5.6) positive for BAC 375G12 proximal end probe were obtained and confirmed as true positives by preparing colony blots. An example of a colony blot filter containing PACs 978I1, 1069K21 and 887N4 is shown. The negative control is a PAC clone located on chromosome 1. Each clone was spotted as two dots. The filter was probed with radiolabelled BAC 375G12 end probe. Autoradiography was carried out overnight at -70^{o}C. All the three clones depicted in the filter were positve for the BAC 375G12 end DNA probe.

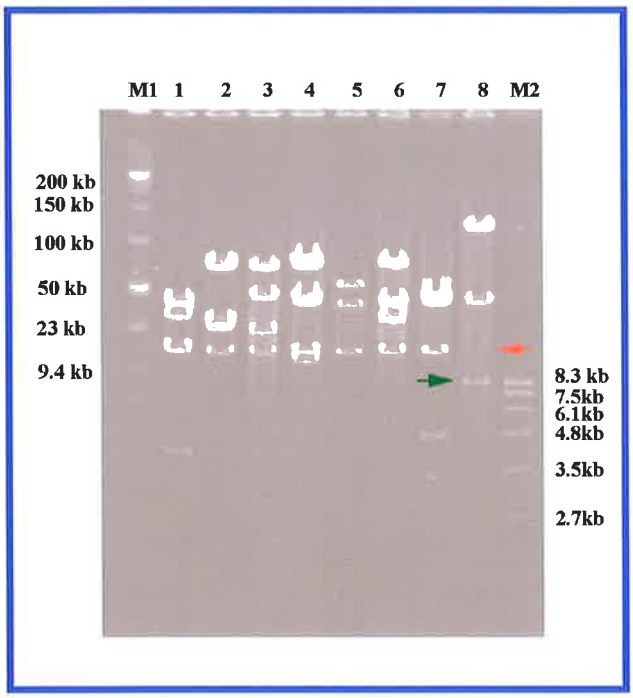


Figure 5.8 Restriction fragment analysis for PAC DNA clones poisitve for BAC 375G12 proximal end probe. Samples: M1) low pulsed-field gel molecular weight marker, 1) PAC 842F20, 2) PAC 978I1, 3) PAC 1054F24, 4) PAC 887N4, 5) PAC 1089J21, 6) PAC 1069K21, 7) PAC 956N2, 8) BAC 375G12, M2) SPP1 marker. The sizes and positions of the marker bands are indicated. BAC 375G12 was used as a positive control in the experiment. The DNA for all the clones was digested using restriction enzymes *Not*I and *Eag*I. Subsequently, the DNA fragments were resolved by PFGE. The conditions for resolution of DNA fragments within the range 5 kb - 150 kb were: gel: 1% agarose gel in 0.5X TBE, running temperature: 14^{O}C, angles of electrodes: 120^{O}, time:15 hours. The restriction fragments obtained for respective clones were added leaving the vector bands to estimate the sizes of the respective DNA clones. The red arrow depicts the 16 kb vector bands of the PAC vector and the green arrow depicts the BAC vector pBACe3.6 specific band. The sizes of respective DNA clones as estimated by PFGE were: PAC 842F20, 80.8 kb; PAC 978I1, 92 kb; PAC 1054F24, 129.8 kb; PAC 887N4, 119 kb; PAC 1089J21, 94 kb; PAC 1069K21, 137 kb; PAC 956N2, 47.6 kb; BAC 375G12, 162 kb.

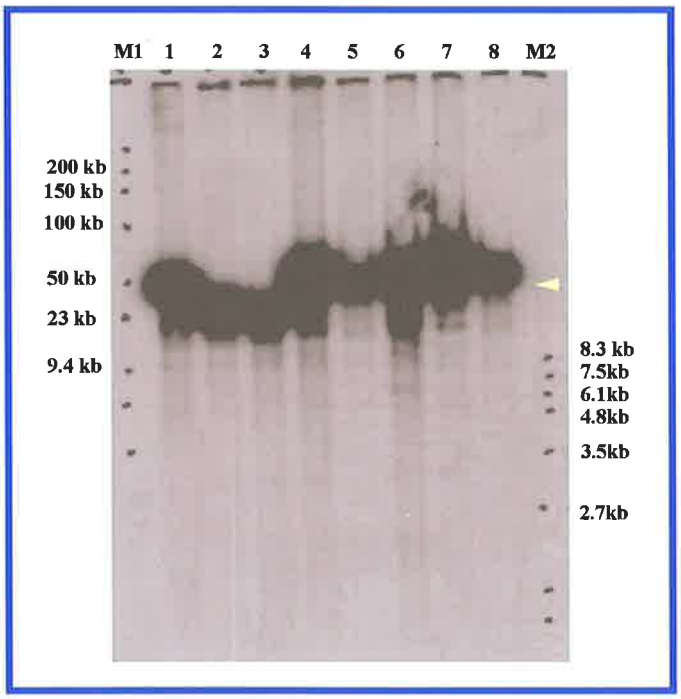


Figure 5.9 Southern hybridisation of PFGE filter containing PAC DNA clones poisitve for BAC 375G12 proximal end probe. Samples: M1) low pulsed-field gel molecular weight marker, 1) PAC 842F20, 2) 978I1, 3) 1054F24, 4) 887N4, 5) 1089J21, 6) 1069K21, 7) 956N2, 8) BAC 375G12, M2) SPP1 marker. The sizes and positions of the marker bands are indicated. BAC 375G12 was used as a positive control in the experiment. The DNA fragments were resolved by PFGE and subsequently, the PFGE filter was made. The filter was probed with PCR product generated using BAC 375G12 proximal end primer pair. Autoradiography was carried out at room temperature for 2 hours to visualise the hybridsed bands, which are depicted by an arrow. The results of hybridisation are summarised in Figure 5.10.

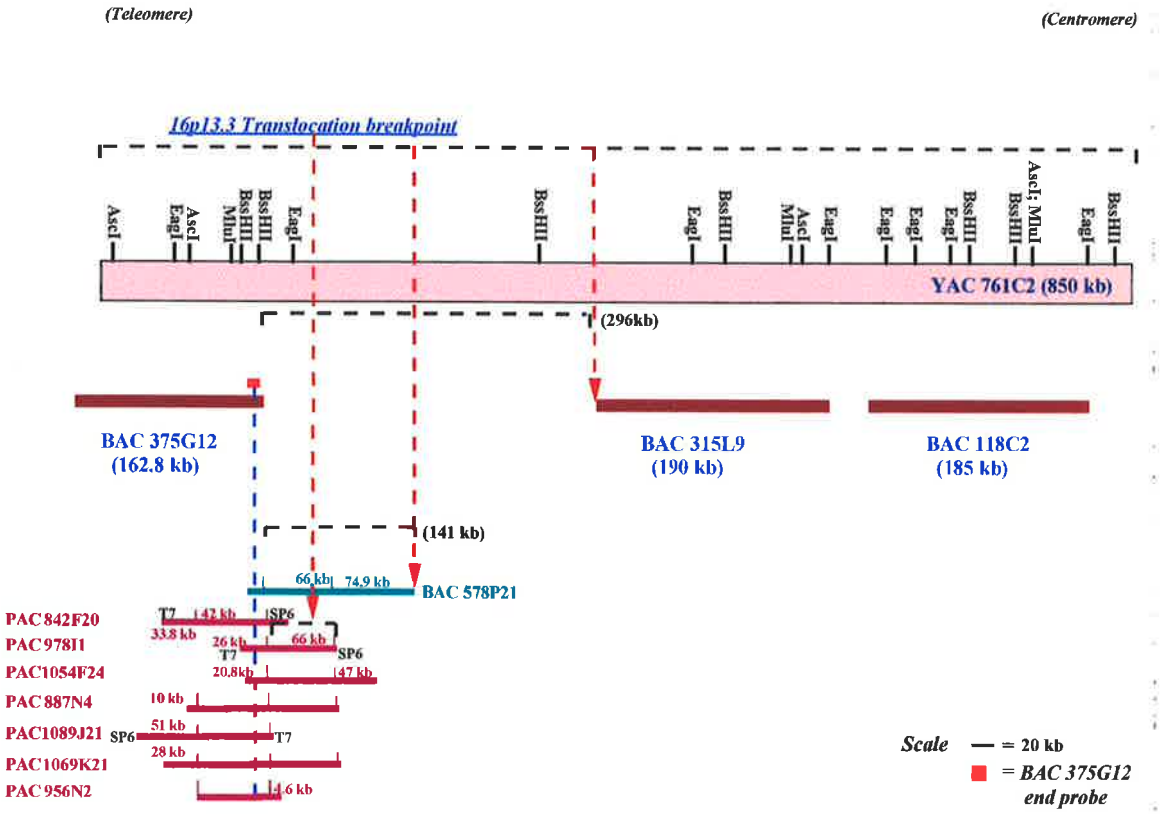


Figure 5.10 Positional cloning of the 16p13.3 breakpoint of the t(1;16) associated with severe mental retardation. The results from Figure 4.4 (YAC restriction map) are also merged for the purpose of Figure preparation. The YAC map data presented in Chapter 4 narrowed the 16p13.3 breakpoint interval in the 296 kb gap between the BACs 375G12 and 315L9, which is shown by a horizontal black dashed line and a red dashed line with an arrow at the end. The other two red dashed lines with arrows at end depict the distance of 16p13.3 translocation breakpoint interval narrowed by the clones 578P21 and 978I1 respectively from the 296 kb gap between the BACs 375G12 and 315L9, during an attempt to clone this breakpoint. The PAC clones identified using BAC 375G12 end probe were digested using restriction enzymes *Not*I and *Eag*I and were arranged into a contig on the basis of the shared common restriction fragments. The sizes of *Not*I/*Eag*I restriction fragments are indicated. The sizes for the same size fragments are depicted only once. The SP6 and T7 ends for the clones whose ends were sequenced are also indicated. The blue dashed line indicates the position of BAC 375G12 end probe on the generated BAC/PAC contig across the 16p13.3. breakpoint. The 16p13.3. breakpoint of the t(1;16) is located within the 66 kb fragment on PAC 978I1.

An example of the colony blot filter hybridised to α^{32}P-dCTP radiolabelled BAC 375G12 proximal end probe is shown in Figure 5.7. All the clones screened (Figure 5.6) were identified to be true positives for the BAC 375G12 proximal end probe.

Subsequently, for all the clones DNA was prepared and digested using restriction enzymes *Not*I and *Eag*I to estimate insert sizes. The digested fragments were resolved using PFGE at the resolution limits ranging from 5 kb to 150 kb, Figure 5.8. The clones were arranged in a possible contig on the basis of their shared restriction fragments. This contig was confirmed by Southern hybridisation analysis to the α^{32}P-dCTP radiolabelled BAC 375G12 proximal end probe (Figure 5.9). Also the SP6 and T7 ends of the PACs 842F20, 978I1, and 1089J21 were sequenced in order to confirm the orientation of the end fragments in the contig. The generated end sequences were aligned using the genomic sequence of the overlapping BAC clone 375G12. The contig spanning the 16p13.3 breakpoint of the t(1;16) is shown in Figure 5.10.

5.3.3.7 FISH analysis of PAC clones 978I1 and 1089J21

PAC DNA clones 978I1 and 1089J21 were used for FISH analysis on the metaphase chromosomes of the patient with t(1;16) *de novo* balanced translocation. PAC 1089J21 was distal to the 16p13.3 breakpoint. The FISH signals for the PAC 978I1 were seen on the normal chromosome 16, and both the derivative chromosomes of the translocation, Figure 5.1 B, indicating that this clone spanned the 16p13.3 breakpoint. Because PAC 1089J21 was localised by FISH distal to the 16p13.3 breakpoint this eliminated the restriction fragment shared with PAC 978I1. Therefore, the breakpoint falls within the 66 kb

restriction fragment of PAC 978I1, reducing the interval from the previous 141 kb (Figure 5.10). Since the size of PAC 978I1, as estimated by PFGE, was only 92 kb in length this was chosen for genomic sequencing.

5.3.4 Sequencing of PAC clone 978I1

The PAC 978I1, which spans the 16p13.3 breakpoint was sequenced to allow the genes and transcripts in the region to be identified. The generated sequence can also be utilised to further localise the 16p13.3 breakpoint in the PAC 978I1. The generated 1.5 X coverage sequence of PAC 978I1 was assembled into 6 blocks. The orientation of the end blocks with respect to BAC 375G12 was determined on the basis of the alignment of the SP6 and T7 ends sequences of PAC 978I1 with the overlapping BAC 375G12 sequence (accession number, AC005751). The sequence generated for PAC 978I1 has been deposited in GenBank under the accession number AC091070.

5.3.4.1 Localisation of 16p13.3 breakpoint on PAC 978I1

The sequence of PAC 978I1, excluding that contained within the sequence of BAC 375G12, was used to design primers for localisation of the 16p13.3 translocation breakpoint of the t(1;16). The primers used are listed in Table 5.2. Each primer pair was localised either distal or proximal to the somatic cell hybrid CY196, which contained the derivative 16 of the t(1;16). The primers from block numbers one to five were localised distal to the 16p13.3 breakpoint contained in CY196 by PCR. While the primers from block number six were localised proximal to this breakpoint, Figure 5.11.

Table 5.2 Primers used to localize the 16p13.3 breakpoint on PAC 978I1

Block No.	Primer name	Sequence (5'–3')		PCR product size (bp)
		Forward Primer	Reverse primer	
Block #1	Block#1/F/R	TGGGCTTTGAGAAATCAAACA	GCCTTTGGGGCTAATTCTCT	202
	Block#1/F2/R2	CAGGTACTAGGGGGTCACCA	AATCAGCAGCCCAGAGAAGA	202
Block #2	Block#2/top/F/R	CCCCAACCCTCCTTTTATTC	GCCTCGATTTCCAAGTGTGT	202
	Block#2/bot/F/R	TCTCCAGTTTGCACCTCCAT	CCCTAAGATCTCCAGCGTCA	187
	Block#2/top/F2/R2	CGGGTTCGACCTATTTTCAA	GCACTTCGTAAGCCACCTTC	192
Block #3	Block#3/top/F/R	GGAGCAGAAAGCCACAGAAC	AGCCCTACACTCCCTCACCT	198
	Block#3/bot/F/R	GAAGTAGGCAACCGCTCAGT	GTGTCTGCGTTTGCTCCTTT	200
Block #4	Block#4/top/F/R	CAGGGGCAGAGAGTTCAGTC	TGTGCCATATGTGGGAAAGA	200
	Block#4/bot/F/R	TTGAGCATGGATGCCATTTA	TGGATTTAGCCCTCAGCAAC	203
	Block#4/bot/F2/R2	TCTGTTAACACCCACCACCA	CAGCTAAGGAGGGAATGTGC	200
Block #5	Block#5/top/F/R	TTGCTGGAATGTTTTGAAGC	TGATGTGAGCTCCTGGATCA	188
	Block#5/bot/F/R	TGCCAGAAGGCCTACAATCT	GCCATGCTCTTCAGGTTTTC	191
	Block#5/bot/F2/R2	GACTCCTCTCATGCCACCAT	ACTCTTAACCTCCCCGGAAA	206
Block #6	Block#6/top/F/R	AACAAGATCCACCATGCACA	CCTGTCTGACGCCTTCAAAT	205
	Block#6/bot/F/R	GTTCAACACTCCGGGAAAAA	CAATGCACGCTTCCCTTATT	203
	Block#6/top/F2/R2	GGTAAGCCAGACCCAAATGA	CTTGGAAACGAAGGTGAAGC	203
	Block#6/c.36/F1/R1	TGATGTGGACTTAATGGCAGA	CCTCCGTTCCTATTTCATGC	167
	Block#6/c.36/F2/R2	AAAGGGAAAAGCATGTGCAG	GAAATGCCCTCATGATCTCC	199
	Block6/c.392/top/F/R	TCCAATACTTGGCTCAGAGGA	TTGGCTTCCAAACCACTCTT	202
	Block6/c.392/bot/F/R	GGGAGTTGGAAAGTGTTGGA	CCACGACTCAGGGAGTAAGC	219
	Block#6/c.301/top/F/R	TGTCGTTGACCATTTTACGG	TGCAGCAACTATTGCCTTTTT	209

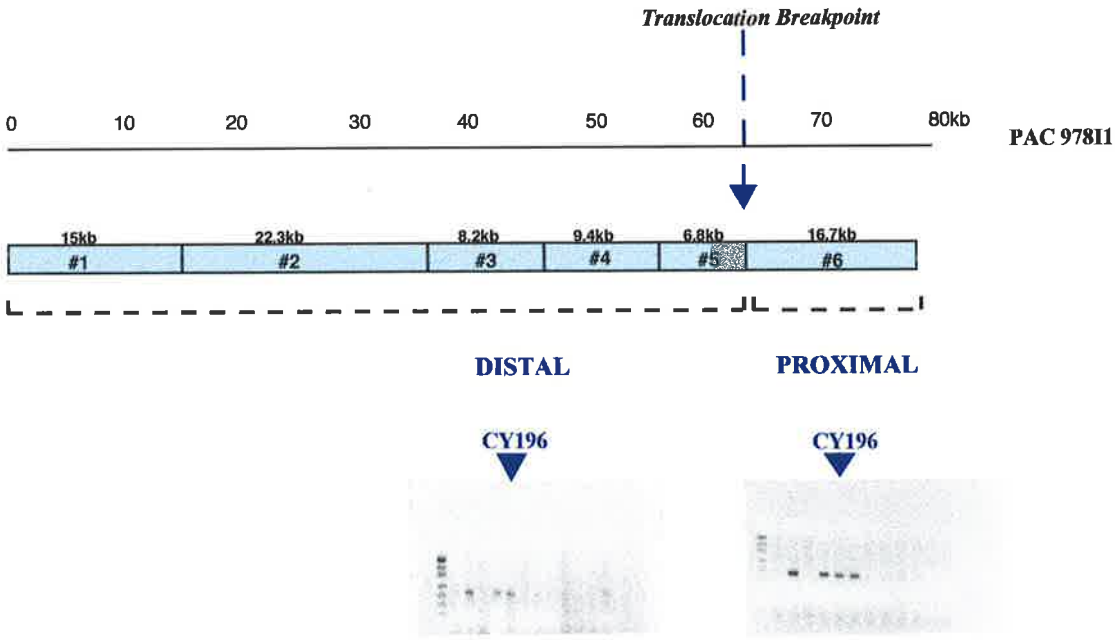


Figure 5.11 Localizaton of the 16p13.3 breakpoint on PAC 978I1 by PCR analysis of the somatic cell hybrids at 16p13.3. The primers were mapped distal or proximal with respect to the CY196 somatic cell hybrid that contained the der(16) of the t(1;16). The 16p13.3 breakpoint was localized between the blocks 5 and 6 of PAC 978I1. The sequence of the PAC 978I1 excluding that contained within the sequence of BAC 375G12, which was distal to the 16p13.3 breakpoint was used for the mapping of 16p13.3 breakpoint of the t(1;16).

Further combinations of the forward and reverse primers designed from the extreme ends of respective blocks, 1-5, Table 5.2, were used with the closest proximal primer pair (Block#6/c.36/R1, Table 5.2) from block number 6 to PCR amplify the breakpoint junction fragment from PAC 978I1. A 1 kb PCR product was amplified with the primers notated as Block#5/bot/R (Table 5.2), designed from block number 5 and the proximal primer pair (Block#6/c.36/R1) from block number 6. None of the other primer combinations yielded any product. Subsequently, the PCR product was purified and sequenced using respective primer pairs used to generate the product. The distal and proximal sequence generated matched the sequences of block numbers five and six, indicating it to be a specific product. It was determined that this sequence was highly repetitive and contained Alu sequences however it was not possible to completely close the gap by standard big dye cycle sequencing chemistry (section 2.3.12.3). The 16p13.3 breakpoint was localised to a 1 kb region between the blocks 5 and 6 on the PAC 978I1 sequence (Figure 5.11).

5.3.4.2 Prediction of ESTs and exons on PAC 978I1 sequence and expression analysis using these predicted ESTs and exons.

The repeats contained within the PAC 978I1 sequence were masked and the sequence initially analysed through dbEST for the detection of matching ESTs using the BLASTN algorithm at NCBI. Very few matching ESTs were detected, indicating that this region was not gene rich. The following ESTs matched the PAC 978I1 sequence: AI243740, AI243860, AI124802, T05472, AI478209 and R88001 (Figure 5.12 and Table 5.3).

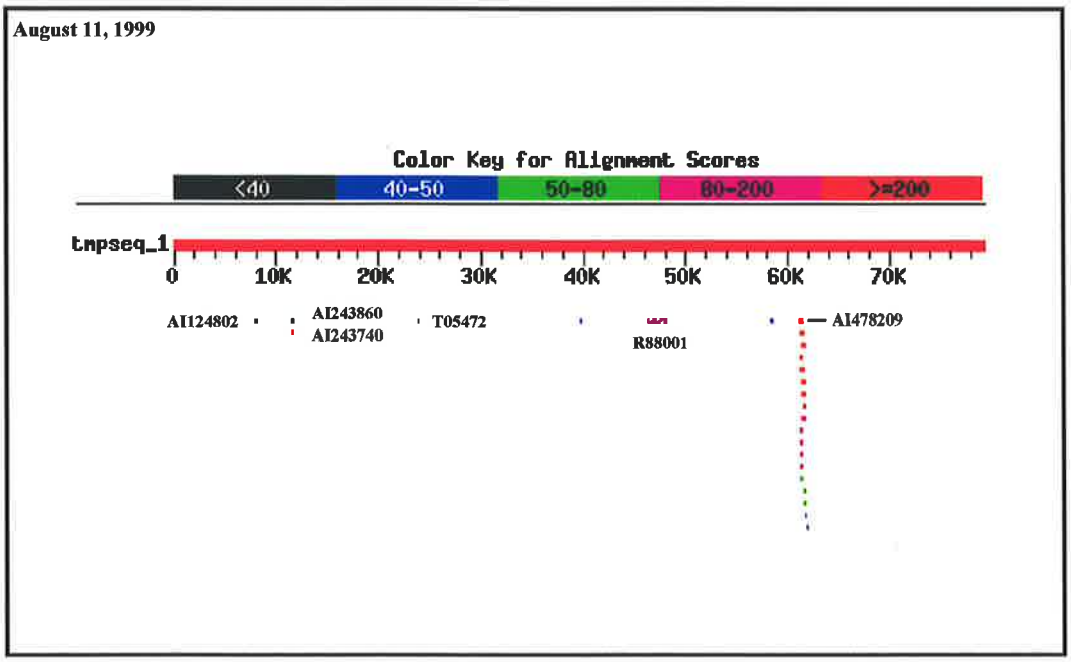


Figure 5.12 dbEST analysis of PAC 978I1 sequence as of August 11, 1999. The analysis identified very few EST matches indicating the region to be gene poor. Only the ESTs used to perform expression studies are indicated in the Figure.

Table 5.3 dbEST analysis of the PAC 978I1 and BAC 375G12 sequences as on August 11, 1999 and March 21, 2000 respectively predicted following ESTs. The information on corresponding cDNA clones, library source, restriction enzymes used to isolate the inserts, and insert sizes is also depicted.

DNA clone	ESTs	CDNA clones	Library	Enzymes	Insert size (bp)
PAC 978I1					
	AI243740	Image:1854113	Soares_NFL_T_ GBC_S1	*NotI*/ *Eco*RI	588
	AI243860	Image:1854191	-do-	*NotI*/ *Eco*RI	588
	AI124802	Image:1539553	Johnston frontal cortex	*Eco*RI/ *Xho*I	1900
	T05472	HFBCZ23	Fetal brain	*Eco*RI/ *Xho*I	1390
	AI478209	Image:2161549	NCI_CGAP_Kid11	-	-
	R88001	Image:165868	Soares adult brain N2b4HB55Y	-	-
BAC 375G12					
	AI902377	BT003	Breast	-	-
	AI207874	Image:1684805	Stratagene schizo Brain S 11	-	-
	M78059	HHCPA67	Hippocampus	-	-
	AL119749	DKFZp761C1924	hamy2	-	-

-do-: same library source, -: the inserts of these clones were not isolated for the expression analysis study as these clones either contained the repeat sequences or were not received from the culture collection centre.

These EST matches were 98-99% identical to the PAC 978I1 sequence except that of the ESTs R88001, and AI478209 (from one of the large group matching PAC 978I1 sequence), which were only 93% and 86% identical to the genomic sequence respectively.

ESTs AI243740 and AI243860

These two ESTs overlapped and formed part of the Unigene cluster number Hs (Homo sapiens) 149074 and were therefore selected in preference to determine possible expression. Both originated from the same library, Soares_NFL_T_GBC_S1, which is pooled from three different normalised libraries, fetal lung NbHL19W, testis NHT, and B-cell NCI_CGAP_GCB1. Only the 3' end sequences were available in the dbEST database therefore the corresponding cDNA clones (Table 5.3) were purchased to allow the sequences of the inserts to be determined. The cDNA inserts were isolated using the respective enzymes (Table 5.3) and their sizes were estimated. Both the cDNA clones had the insert sizes of 588 bp. Therefore, the insert for one of the cDNA clones, Image:1854113, was sequenced. The sequence generated was aligned with the PAC 978I1 sequence and was contiguous with no evidence of any exon/intron boundaries and an absence of a poly A signal.

Further the α^{32}P-dCTP radiolabelled cDNA insert (cDNA clone Image:1854113) when hybridised to the RNA Master blot and Multiple Tissue Northern blot did not identify any corresponding transcript. The Multiple Tissue Northern Blot contained poly (A^+) mRNA from various adult human tissues, including heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas.

EST AI124802

The 3' end sequence (419 bp) of this EST was only available in the dbEST database and this contained a MER20.b2 repetitive element. Therefore, the corresponding cDNA clone was obtained to generate additional sequence in order to design primers without the repeat sequence for use in expression studies.

Table 5.4 Primers used for the RT-PCR analysis.

Primer name	Primer sequence 5'-3'		PCR product size (bp)
	Forward primer	Reverse primer	
AI124802 F1/R1	TGC TTA AAA CCC ACC ACA GC	TCT GCT CCA CCT TTG TTA ATG	1290
AI124802 nor/1F/1R	TCG TTT CTA GGC AGG GAA TG	CCA TTA ACA AAG GTG GAG CA	388
AI478209 F/R	CGG CCT GCA AGA TCC TGT C	ACC TGG GAT GTT TCC CTG AC	233
RT-PCR NIX #1	TGC CAG AAG GCC TAC AAT CT	GCC ATG CTC TTC AGG TTT TC	191
RT-PCR NIX #2	ATC CCC TTT CTC CCC TCT GT	TCT GGT ACC AAA GGG TGT CC	139

The cDNA insert was isolated using restriction enzymes (Table 5.3) and the size of the clone was estimated to be 1.9 kb. The 5' and 3' end sequence was generated. The primers AI124802 F1 and R1 (Table 5.4) were designed from the generated 5' and 3' end sequences to obtain the remaining sequence of the 1.9 kb insert. This cDNA insert sequence was aligned with the PAC 978I1 sequence, which was also contiguous with PAC sequence with no evidence of any exon/intron boundary and poly A signal.

To detect if this EST corresponds to an exon, RT-PCR analysis was done using the primers AI124802 F1 and R1, and AI124802nor/1F and 1R (Table 5.4) and

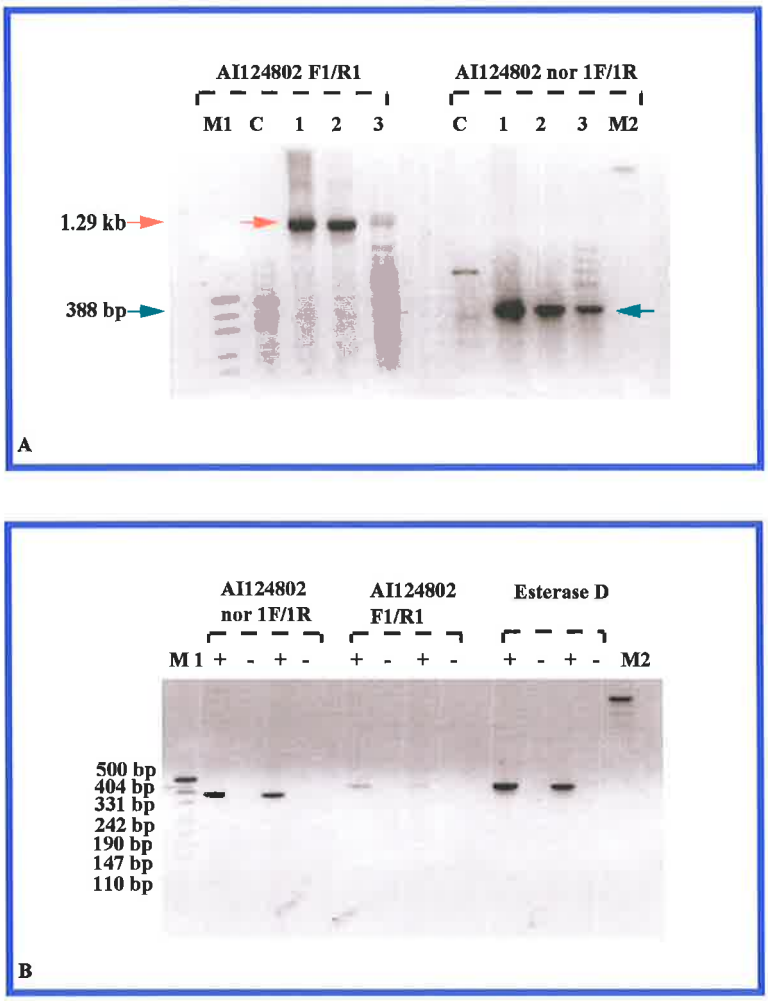


Figure 5.13 PCR and RT PCR analysis of EST AI124802. **(A)** PCR amplification of cDNA and genomic DNA using primer pairs designed from AI124802 sequence. Both the primer pairs amplified the same size products from cDNA and genomic DNA. The red arrow indicates the product obtianed using AI124802 F1/R1 primer pair and green arrow indicates the product obtianed using AI124802 nor F1/R1 primer pair. Samples: M) pUC 19, C) no DNA control, 1) cDNA AI124802, 2) PAC 978I1 DNA, 3) blood bank DNA, M2) SPP1 marker. **(B)** The primer pairs AI124802 F1/R1 and AI124802 nor F1/R1 were also used for RT PCR analysis. The fetal brain poly mRNA was reverse transcribed using both oligo dT and random hexamers. + represents lanes with reverse transcriptase enzyme and - represents the lanes without reverse transcriptase enzyme. Esterase D primer pairs were used as the positve control in the experiment. The sizes for pUC19 marker bands are indicated.

fetal brain poly A mRNA. Esterase D primer pairs (section 2.3.8), which only amplify a product from cDNA were used as a positive control to check the success of the cDNA synthesis. The primer pair AI124802 F1 and R1 amplified a faint product of approximately 435 bp (Figure 5.13 B). However, this faint RT-PCR product was more likely to be a genomic artefact because the primer pair AI124802 F1 and R1 amplified the same size PCR products (1.29 kb) from the cDNA clone and genomic DNA (Figure 5.13 A). The faint 435 bp RT-PCR product was therefore cloned and subsequently sequenced. The sequence generated did not match the cDNA and PAC 978I1 sequence, which revealed that it was a non-specific product. The second set of the primers AI124802nor/1F and 1R amplified a 388 bp RT-PCR product (Figure 5.13 B). Because this RT-PCR product was of similar size to the product obtained by PCR amplification of cDNA clone and genomic DNA (Figure 5.13 A) it was purified and used to hybridise to a Multiple Tissue Northern blot. However, the 388 bp RT-PCR product failed to identify any transcript, which indicated that this product could possibly be the amplified genomic DNA present in the RNA sample or a non-specific product. It was unlikely that the corresponding transcript was expressed in a tissue not present on the Northern because the EST AI124802 was derived from brain tissue and this was present on the Northern blot.

In the absence of the hybridisation signals being detected by the DNA probes corresponding to the ESTs AI243740 and AI124802, the Northern membrane was probed with a positive control, which detected specific hybridisation signals.

The sequences obtained for the cDNA clones Image 1854113 and 1539553 corresponding to the ESTs AI243740 and AI124802 respectively (Table 5.3)

completely matched the PAC 978I1 sequence. The sequence for PAC 978I1 has been deposited in GenBank under the accession number AC091070. Therefore, the sequences for cDNA clones Image 1854113 and 1539553 are not separately presented in this thesis.

EST T05472

Only 219 bp of the sequence was available for this EST in the database. Therefore, the corresponding cDNA clone (Table 5.3) was obtained to generate further sequence of the cDNA insert. The insert size was 1.39 kb (Table 5.3). The 5' and 3' ends of the clone were sequenced but the sequence generated did not match with the PAC 978I1 sequence. This indicated that the culture collection centre sent the incorrect clone or the entries in the databases are incorrect. Therefore, further RT-PCR analysis was not performed with this EST.

AI478209

The EST AI478209 originated from a library NCI_CGAP_Kid11 (Table 5.2) derived from kidney tissue. This EST was only 86% identical to the PAC 978I1 sequence. Therefore, in order to check if it is a true positive match for PAC 978I1 sequence, the forward and reverse primers AI478209 F and R (Table 5.4) were designed from the sequenced regions of this EST, which were not completely identical to PAC 978I1 sequence. Using this primer pair no PCR product was amplified from the PAC 978I1 DNA, which indicated that this EST was a false positive match to the PAC 978I1 sequence. This EST was discarded from further analysis.

R88001

This EST was located at the 5' end of the cDNA clone Image: 165868. 487 bp of the 5' end sequence was available in the dbEST database that was only 93% identical to the PAC 978I1 sequence with 6% gaps in the sequence alignment. The 3' end sequence for this clone was not available in the database. Therefore, the corresponding cDNA clone was ordered to generate the insert sequence and to first determine if this EST was a true positive match on the PAC 978I1 sequence to be able to perform further expression analysis. However, it was not possible to work with this EST because the corresponding cDNA clone (Table 5.3) was dead in the culture collection centre and therefore was not received.

Expression analysis with the ESTs, which matched the PAC 978I1 sequence did not identify any potential transcript on PAC 978I1. Therefore, the computer prediction of exons was also performed using the NIX program (UK HGMP Resource Centre) (Figure 5.14), which enables the results from several different gene predicting programs to be displayed simultaneously on the genomic sequence. This analysis also indicated that the region surrounding the breakpoint is gene poor with very few matches to expressed sequences or computer predicted exons. Most of the exons predicted were present as singletons, that is, were predicted by only one particular program and not by other programs. Therefore, these regions were less likely to be true exons. However, there were two regions, which are indicated in the Figure 5.14, where exons were predicted by four different programs in the same region of the genomic sequence.

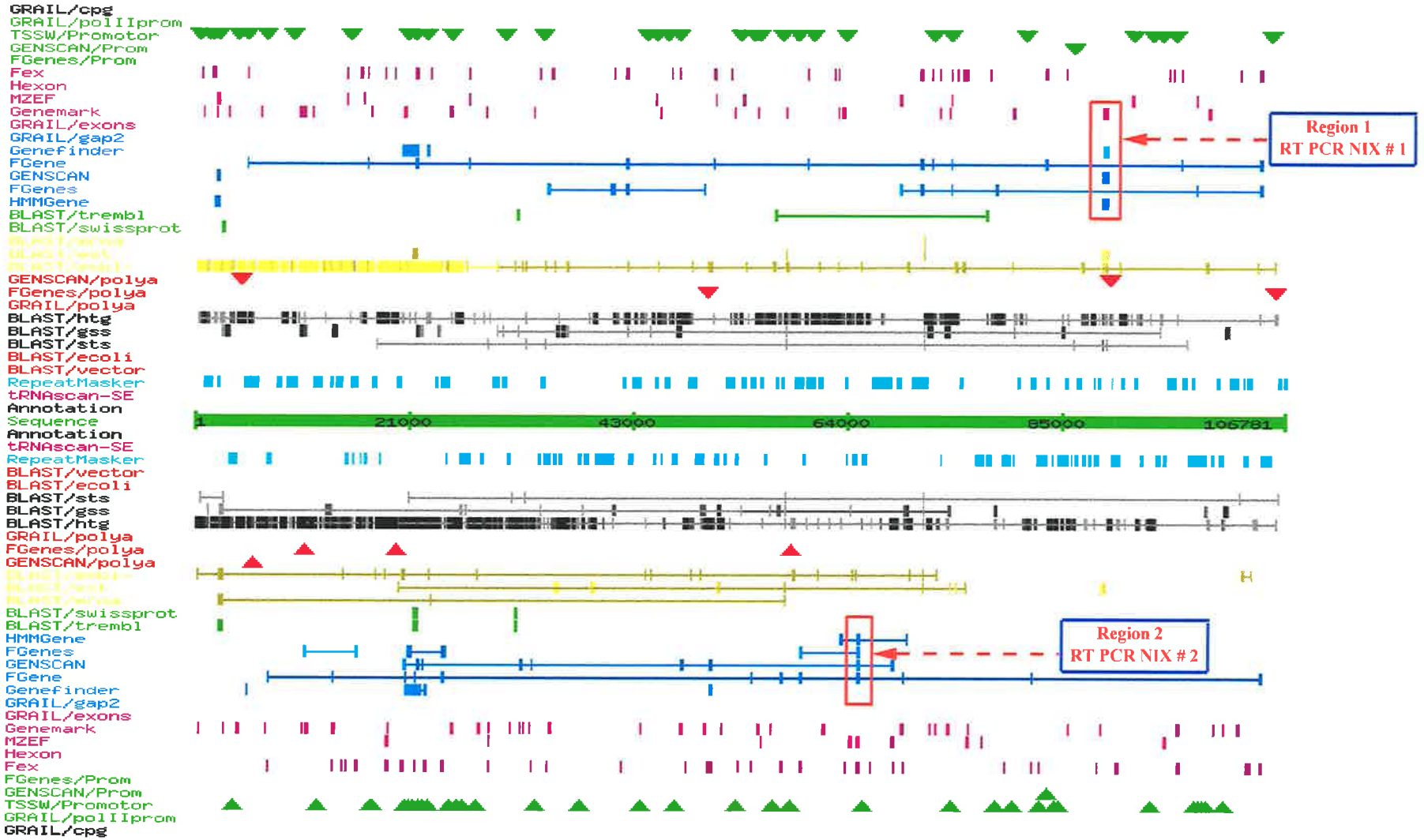


Figure 5.14 NIX (UK HGMP Resource Centre) analysis of PAC 978I1 sequence as of November 1, 1999. This program enables different gene predicting programs to run simultaneously on the genomic sequence. Most of the predicted exons were present as singletons. There were two genomic regions predicted as exons by four different programs, which were used for the expresssion analyses. These two genomic regions in the Figure are boxed. The primer pairs designed from these two regions for expression studies are indicated.

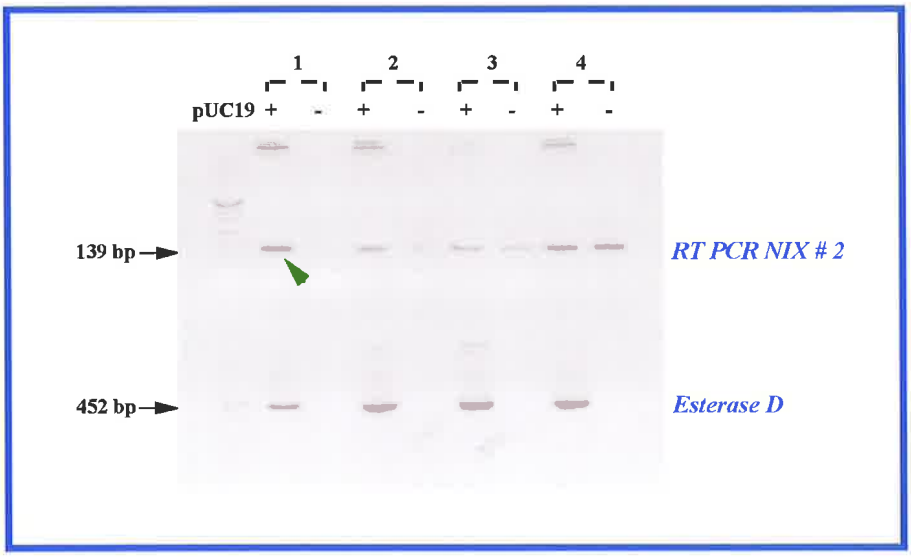


Figure 5.15 RT PCR analysis of NIX predicted exon, designated as RT PCR NIX # 2. Samples: 1) fetal brain poly A mRNA, 2) total RNA from fetal spleen, 3) total RNA from fetal kidney, 4) total RNA from fetal lung. + indicates lanes with reverse transcriptase enzyme and - indicates lanes without reverse transcriptase enzyme. The product sizes are indicated. Esterase D was used as a positive control in the experiment. The RT PCR product amplified from fetal brain poly A mRNA, depicted by an arrow was purified and used as probe on Northern blot, that failed to identify any transcript.

Primers (Table 5.4) were designed from these regions for RT-PCR analysis of fetal brain poly A mRNA, and total RNA samples from fetal spleen, fetal kidney and fetal lung. Esterase D primers were included as positive control in the experiment. The RT-PCR products were only amplified by the primer pair RT-PCR NIX #2, the results of which are shown in Figure 5.15. However, for these primers the same sized bands were also seen in the negative control (lanes without the reverse transcriptase enzyme) lanes containing total RNA from fetal spleen, fetal kidney and fetal lung. These products were likely to be amplified from the contaminating genomic DNA present in the RNA samples because the esterase D control primers did not amplify any products in the negative control lanes of these samples. This only exception was the fetal brain poly A mRNA where no band was seen in the corresponding negative control. Therefore, this 139 bp RT-PCR product was purified (section 2.3.2.2.1) and used as a probe to hybridise to a Northern blot for transcript identification. This probe failed to identify any corresponding transcript. The PCR amplification of this product could be a result of genomic artefact or the transcript might be expressed in a tissue not present on the Northern membrane used.

The ESTs matches identified by dbEST BLAST analysis (August 11, 1999) of PAC 978I1 sequence and the computer predicted exons identified by NIX analyses (November 1, 1999) of PAC 978I1 sequence did not identify any transcript by RT-PCR or Northern analyses. PROSITE and ProDom databases searches using the entire PAC 978I1 sequence showed no significant homology to known proteins and domains. Therefore, the sequence of the overlapping adjacent BAC 375G12 was also analysed by BLAST search analysis using dbEST database to detect matching ESTs near the 16p13.3 breakpoint. The BAC 375G12

contained 43% repeat sequences and the dbEST BLAST search results as of March 21, 2000 detected only four EST matches on the BAC 375G12 sequence (Table 5.3). The corresponding cDNA clones were ordered for each of these ESTs but only two clones Image:1684805 (EST AI207874) and DKFZp761C1924 (EST AL119749) were available. The end sequences of the inserts were generated. These sequences contained Alu repeats, and further did not identify any intron boundary on alignment with the genomic sequence. Because these ESTs were present as singletons, had repeat elements and no evidence of any intron/exon boundaries it was unlikely that these ESTs would correspond to any transcript except if they were a part of the UTR region of a gene. Therefore, further steps towards expression studies using these ESTs were not performed.

5.4 Discussion

A positional cloning strategy was used to define the 16p13.3 translocation breakpoint in the patient with a *de novo* balanced translocation t(1;16) associated with severe mental retardation. Chromosomal rearrangements have greatly assisted the positional cloning of disease causing genes (Collins, 1995). Since the chromosome 1 breakpoint was shown to be within the 1q heterochromatin, the search for the gene responsible for the patient's phenotype was focussed on the chromosome 16 breakpoint.

The long range YAC map data presented and discussed in Chapter 4 (Figure 4.7) narrowed the 16p13.3 breakpoint of the t(1;16) to a 296 kb distance between the BACs 375G12 and 315L9. To further refine the breakpoint interval, the BAC 375G12 sequence was used to design end probes to screen a PAC library and for BLAST search analysis to identify any overlapping DNA clones that would lie in

the 296 kb gap. Firstly, the orientation of the BAC 375G12 with respect to the YAC 761C2 was determined by mapping the end probes generated from the BAC 375G12 with respect to the YAC 761C2 and to somatic cell hybrids at 16p13.3 including the hybrid CY196, which contained the der(16) of t(1;16). The proximal end probe of the BAC 375G12 was contained within YAC 761C2.

Database searching using the BAC 375G12 sequence identified the T7 end of BAC 578P21. Subsequent, PCR and Southern analyses using the end probes generated from the BAC 375G12 sequence confirmed that the BAC 578P21 overlapped with the proximal end of BAC 375G12. Restriction fragment analysis using PFGE estimated the size of BAC 578P21 to be 147 kb. This BAC further reduced the 296 kb gap between the BACs 375G12 and 315L9 to 141 kb. FISH analysis with this clone on the metaphase chromosomes of the patient with t(1;16) indicated that this clone spans the breakpoint of the translocation at 16p13.3.

However, to further narrow the region of 16p13.3 breakpoint to a smaller DNA clone, which could be more economically sequenced, a PAC contig was also constructed across the CY196 breakpoint. For this, the end probe of BAC 375G12, which was contained within the YAC 761C2 was used to screen a human PAC (RPCI5) library. The positive DNA clones identified were arranged in a contig on the basis of the similar size restriction fragments and hybridisation of these clones to the BAC 375G12 end probe.

The PAC clones 978I1 and 1089J21 were used for the FISH analysis on the metaphase chromosomes of the patient with t(1;16). PAC 1089J21 was distal to the 16p13.3 breakpoint, where as the FISH signals for the PAC 978I1 revealed

that this clone spanned the 16p13.3 translocation breakpoint of t(1;16). This enabled the 16p13.3 breakpoint interval to be reduced to 66 kb. DNA of this spanning PAC was sequenced. The PAC sequence with 1.5 X coverage was assembled into 6 blocks. Analysis of this sequence by PCR localised the 16p13.3 translocation breakpoint to a 1 kb region between the blocks 5 and 6 on the PAC 978I1 sequence.

Analysis of the PAC 978I1 and BAC 375G12 sequence using dbEST and the NIX program indicated that the region surrounding the breakpoint is highly repetitive with very few matches to expressed sequences (ESTs) or computer predicted exons. These ESTs and exons were present as singletons, none of these ESTs had a poly A signal, no homology was observed to other known proteins and there was no evidence of any exon/intron boundaries. Therefore, it was concluded that it was unlikely that these transcripts represented transcribed genes. Furthermore, expression studies with some of these EST matches and predicted exons failed to identify any evidence of an expressed transcript. This suggested that the chromosomal rearrangement either resides in a large intron of a causative gene, the observed phenotype is the result of a position effect of the chromosome 1 heterochromatin or the association of translocation and clinical abnormalities is fortuitous.

The physical mapping data presented in Chapter 3 had identified C16*orf*5, a proline rich transcript highly expressed in the brain, which is the closest identified gene distal to the 16p13.3 breakpoint of the t(1;16). Therefore, the position effect of the adjacent translocated chromosome 1 heterochromatin on the expression of C16*orf*5 gene was determined, which is presented in Chapter 7. Juxtaposition of a

gene to a region of heterochromatin through chromosomal rearrangements can result in the alteration of gene expression (Karpen, 1994).

Subsequent to this work being completed a dbEST BLAST search using the BAC 375G12 sequence performed on November 11, 2000 revealed that the BAC 375G12 contains the 5' UTR of the novel protein A2BP1 (ataxin–2 binding protein 1). A2BP1 was recently identified and sequenced (Shibata et al., 2000). The 16p13.3 breakpoint of the t(1;16) characterized in this Chapter was found to be present in an intron of this gene. This explains the failure of transcript identification in this region. The physical mapping data of the A2BP1 gene with respect to the 16p13.3 breakpoints of the t(1;16) and t(14;16) is presented in Chapter 7.

Positional Cloning of the t(14;16) associated with Mild Mental Retardation and Epilepsy

Chapter 6

6.1 Introduction

A male patient with a balanced *de novo* translocation t(14;16)(q32;p13.3) associated with mild mental retardation and grand mal seizures was reported in the Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, Adelaide. It is proposed that the chromosomal translocation in this patient is related to a disruption of a gene at or near the translocation breakpoint and this is responsible for the observed phenotype.

Mental retardation and/or epilepsy account for the majority of childhood brain disorders. Mental retardation (MR) is described as significantly sub-average intellectual functioning and limitations in adaptive skills, with onset before 18 years of age (American Psychiatric Association, 1995). A seizure is a paroxysmal and transitory disturbance of brain function that develops suddenly and terminates spontaneously. It is physiologically characterized by abnormal, excessive and self-terminating discharges from neurons (Zara et al., 1995). Approximately 2%-3% of the human population has either learning or behavior disabilities with an intelligence quotient (IQ) less than 70 and is subdivided into ranges: borderline (~70), mild (50-69), moderate (35-49), severe (20-34), and profound (<19) (McLaren and Bryson, 1987; Gecz and Mulley, 2000). Extensive population based studies reported by Steffenburg et al., (1995) and Stromme and Hagberg (2000) into the causes of MR suggests that epilepsy is one of the common neurological impairments occurring in a population with MR. In a population-based study of Norwegian children with MR conducted by Stromme and Hagberg (2000), MR in 20% of the children was associated with epilepsy. This proportion was found to be (33%) in children with severe mental retardation and 9% in children with mild mental retardation. Epilepsy comprises a heterogeneous group

of seizure disorders, affecting 3% of the population during life (Hauser et al., 1996; Sander et al., 2000).

The hunt for a disease gene was focused on chromosome 16. The two translocations t(1;16) and t(14;16) characterized in the present study were mapped to the same physical interval on 16p13.3 and it was hypothesized that MR and epilepsy in the two patients was due to disruption of the same gene at 16p13.3.

In order to refine the 16p13.3 translocation breakpoint of the t(14;16), a BAC/PAC contig was constructed across the breakpoint using the closest flanking probes. FISH analysis with these clones to metaphase chromosomes of the patient was done to identify a clone that spanned the breakpoint. The sequence of this spanning clone was analyzed for the identification of candidate genes.

This Chapter presents the molecular characterization of the t(14;16)(q32;p13.3) *de novo* balanced translocation associated with mild mental retardation and grand mal seizures.

6.2 Methods

The majority of methods used in this chapter were the general molecular genetic techniques of DNA analysis, which are described in detail in Chapter 2.

6.2.1 Somatic cell hybrid analysis

The somatic cell hybrid analysis was performed as described earlier in section 2.3.6. The primer pairs used in this Chapter were designed as described in section 2.3.5.1.

6.2.2 Screening of PAC/BAC libraries

Screening of three different high-density PAC/BAC libraries was performed as described in section 2.3.7. The following human PAC/BAC libraries, PAC library RPC14, PAC library RPCI5 and BAC library RPCI-11 Segment 4 (PAC/BAC Resources, Buffalo, New York) were screened using the closest probes flanking the 16p13.3 breakpoint. To ensure that cross hybridization was absent, the probes were initially hybridized to the PAC vector filters, prepared (section 2.3.7.4.1) by digesting the PAC vector DNA with restriction enzyme *Bam*HI, and human genomic filters, prepared (section 2.3.4.7.2) by digesting the genomic DNA with restriction enzyme *Pst*I.

6.2.3 Identification of PAC/BAC clones

Individual PAC/ BAC DNA clones identified from screening the high-density PAC/BAC filters (6.2.2) were obtained from PAC/BAC Resources, Buffalo, New York. Additional BAC clones were identified by BLAST search analysis and were obtained from Research Genetics, USA. The BAC clones 192K18 and 26O3 were kindly provided by Dr. Norman Doggett, Los Alamos National laboratory, New Mexico. The BAC clones located from 16p12.2 to 16p11.2 were from the physical map of Cao et al., (1999), which were kindly provided by Dr. Mei Wang (California Institute of Technology, Pasadena, USA).

6.2.4 DNA isolation

Clones were grown at 37°C in LB (Luria Bertoni) medium (section 2.2.6.1) containing 50 ug/ml of kanamycin, in the case of PACs, or 30 ug/ml of chloramphenicol for BACs. DNA was isolated using Qiagen 100 columns (section 2.3.1.3). The clone inserts were sized and restriction mapped using pulsed-field

gel electrophoresis (PFGE) (section 2.3.4.4.3) and the infrequently cutting restriction enzymes *Not*I and *Eag*I (section 2.3.4.1). The gels were then transferred to Hybond-N^+ nylon membranes as described in section 2.3.4.7.1. The radiolabelling of probe, membrane hybridization and washing was done as per the methods described in sections 2.3.4.

6.2.5 Fluorescent in situ hybridization (FISH)

FISH analysis was performed by Helen Eyre (Department of Cytogenetics and Molecular Genetics, WCH, Adelaide) as described in section 2.3.11. The DNA clones identified using the probes located distal and proximal to the 16p13.3 breakpoint were analyzed by FISH to the metaphase chromosomes of the patient with the t(14;16) in order to determine if these clones span the 16p13.3 breakpoint.

FISH analysis was also done using a series of probes located on the short arm of chromosome 16 to resolve the chromosomal rearrangement in this patient. For use as FISH probes, cosmids, PACs and BACs were labelled by nick translation with biotin-14-dATP (GIBCO-BRL). For chromosome painting, commercially prepared whole chromosome paints for chromosome 14 (Cambio, DIG-labelled) and chromosome 16 (Cambio, biotin-labelled) were hybridized to metaphase chromosomes according to the manufacturer's instructions and detected with TRITC-anti DIG for chromosome 14, and avidin-FITC for chromosome 16.

6.2.6 Computational analysis

Searches of the gss and htgs databases were performed using the BLAST algorithim at NCBI with the BAC 441K8 sequence (accession number,

AC007222) for the identification of overlapping DNA clones. A dbEST database analysis of the BAC 545E8 sequence (accession number AC022206) was performed for the identification of ESTs matching the genomic sequence.

6.2.7 Northern analysis

Northern analysis was performed to determine the size and tissue expression of a transcript corresponding to the cDNA clones, Image:44969 and Image:25052, by hybridizing the cDNA inserts to a commercially obtained Northern blot (Clontech, Palo Alto, USA) as described in section 2.3.9.

6.3 Results

6.3.1 Case report of the patient with t(14;16)

Phenotype of the patient

This boy was born after an uneventful pregnancy and labour. During the first three months of life he had feeding difficulties and abdominal pain. He sat at 8 months and walked at 18 months. At 14 years of age his speech was normal and he was in special school being assessed as having mild mental retardation. He had a grand mal seizure at the age of 5 and at 7, and a further two from 7 to 14 years. Short absences were noted at about monthly intervals. Anticonvulsive therapy was initiated at age 7. At the age of 7 he was in the 97^{th} percentile for height, weight and head circumference. Motor abilities were retarded and there were behaviour difficulties. EEG was normal. Clinodactyly of the 5^{th} finger and mild syndactyly of the 2^{nd} and 3^{rd} toes were the only dysmorphisms.

Karyotype of the patient

The classical G-banding of chromosomes from the patient demonstrated a *de novo* translocation between chromosomes 14 and 16, t(14;16)(q32;p13.3), Figure 6.13 a. A mouse/human somatic cell hybrid was generated from the cell line of this patient containing the der(16)t(14;16) designated as CY182 (Callen et al., 1995). Later in the course of this thesis, during an attempt to clone the 16p13.3 breakpoint contained in CY182, a paracentric inversion of a portion of the short arm of 16 was identified with the distal breakpoint coincident with that of the translocation at 16p13.3 and the proximal breakpoint at 16p12.1. This complex three break rearrangement can be notated as t(14;16)(p32;p13.3)inv16(p13.3p12.1).

6.3.2 Cloning of the 16p13.3 breakpoint of t(14;16)

6.3.2.1 Cosmid c62F6

Probes located proximal to the 16p13.3 breakpoint in CY182 were initially used to construct a PAC/BAC contig. Primers (Table 6.1) were obtained for the cosmid c62F6 that was located proximal to the breakpoint in CY182 in the previously constructed chromosome 16 map in the Department of Cytogenetics and Molecular Genetics, WCH, Adelaide. The physical mapping data presented in Chapters 3 and 4 refined the order of chromosome breakpoints of hybrids at 16p13.3 from that reported earlier (pter-23HA, CY182, CY177-CY196, CY197-CY198-CY168) in Callen et al., (1995) to pter-23HA-CY196-CY197-CY182, CY177-CY198, CY168. Therefore the localization of c62F6 was first confirmed by PCR analysis of these somatic cell hybrids. The results of mapping are shown in Figure 6.1.

Figure 6.1 Mapping of c62F6 with respect to the 16p13.3 breakpoint in CY182 using somatic cell hybrids at 16p13.3. Samples: 1) Human DNA, 2) cosmid c62F6 DNA, 3) A9, 4) CY18, 5) 23HA, 6) CY194, 7) CY177, 8) CY182, 9) CY196, 10) CY197, 11) CY198, 12) CY168. The CY18 and A9 somatic cell hybrids were used for determination of the chromosome 16 specific bands and mouse bands respectively. Lane C, represents no DNA (water) control and M is pUC19 marker DNA. The cosmid c62F6 was located on chromosome 16 in the hybrid interval defined by CY182, CY177 and CY198. The size of PCR product is indicated.

The cosmid was localized in the hybrid interval defined by CY182, CY177 and CY198. The results were then confirmed by FISH analysis on metaphase chromosomes of the patient with the t(14;16). This showed that c62F6 was proximal to the 16p13.3 breakpoint of the t(14;16), which was consistent with PCR mapping.

6.3.2.2 Screening of the human PAC library (RPC14)

In order to identify a DNA clone spanning the 16p13.3 breakpoint contained in CY182, the human PAC library RPC14 was screened using the PCR product generated from c62F6. Five clones were initially identified and subsequently confirmed as probable positives by colony blot hybridization using the c62F6 probe. An example of the colony blot filter hybridized to α^{32}P-dCTP radiolabelled c62F6 probe is shown in Figure 6.2 A. Subsequently, hybridization of Southern blots with α^{32}P-dCTP radiolabelled c62F6 probe showed that only three of these clones were true positives, 722G3, 722F24 and 768B15.

PAC DNA was prepared and digested using the restriction enzymes *Not*I and *Eag*I to estimate their insert sizes. The digested fragments were resolved using PFGE at the resolution limits ranging from 5 kb to 150 kb. The restriction fragment patterns obtained are shown in Figure 6.2 B. The PACs were then arranged into a contig on the basis of shared similar size restriction fragments and probing with the α^{32}P-dCTP radiolabelled c62F6 probe, Figures 6.2 B and C.

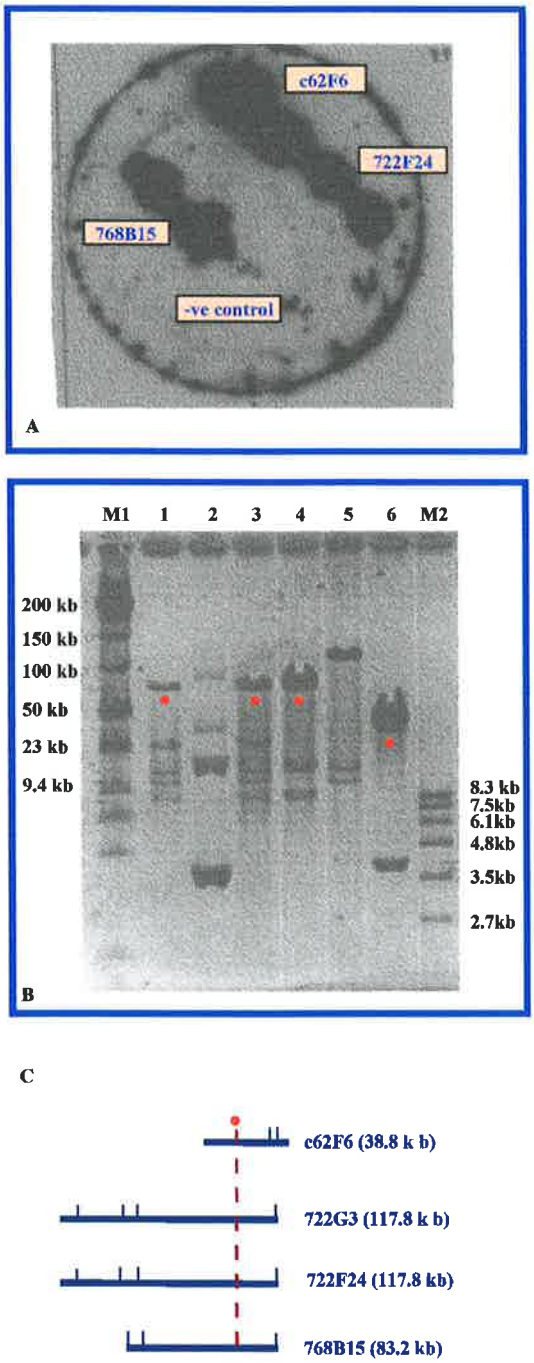
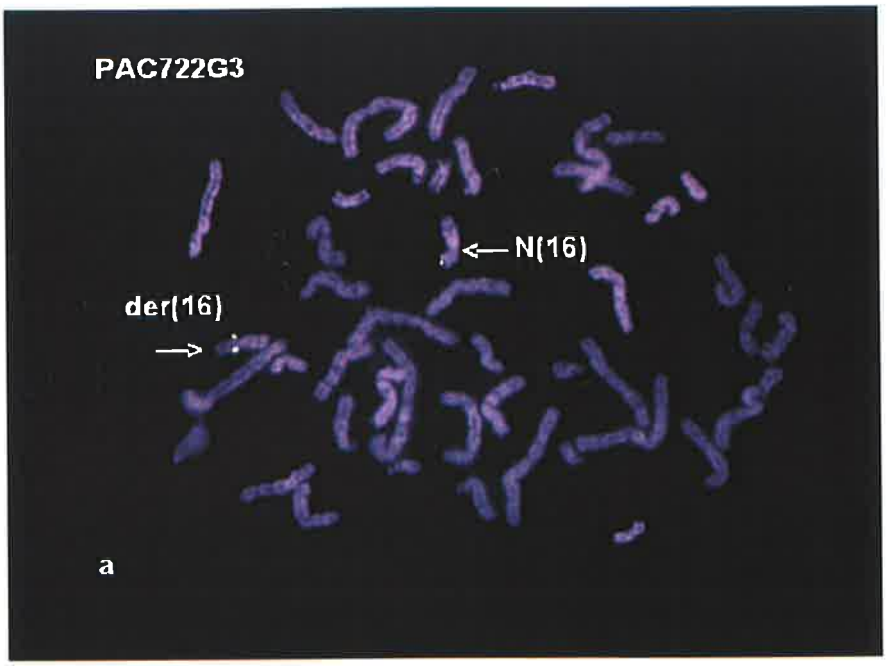
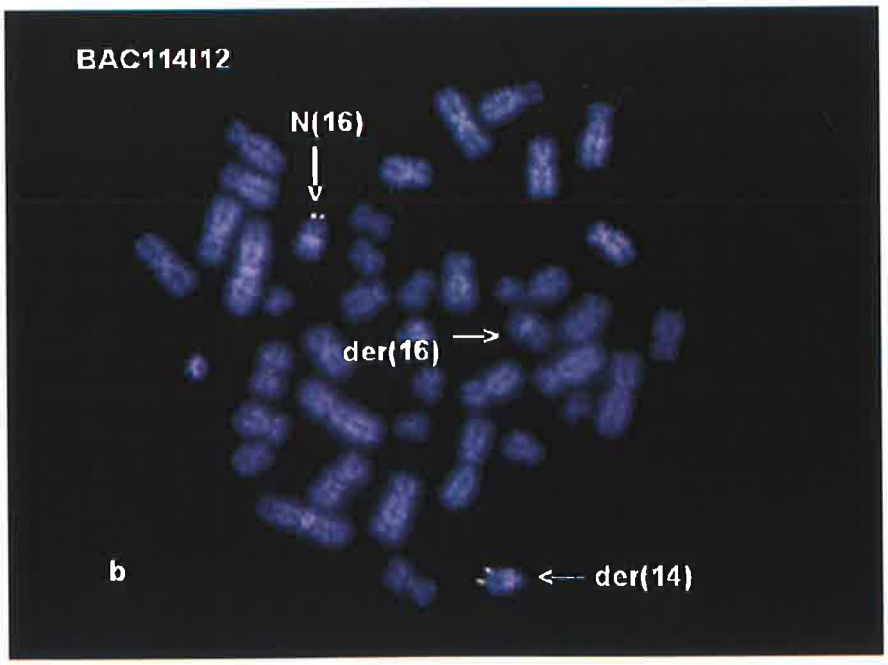


Figure 6.2 Experiments performed to arrange the PAC clones screened from RPC14 human PAC library using c62F6 probe into a possible contig. **(A)** Colony blot hybridisation. The PAC DNA clones positive for c62F6 were obtained and confirmed for true positives by preparing colony blots. An example of the colony blot filter containing PACs 768B15 and 722F24 is shown. c62F6 cosmid was used as a positive control. The negative control is a PAC clone located on chromosome 1. Each clone was spotted as two dots. **(B)** Restriction fragment analysis for PAC DNA clones poisitve for c62F6 probe. Samples: M1) low pulsed-field gel molecular weight marker, 1) PAC 722F4, 2) PAC 814O18, 3) PAC 722G3, 4) PAC 768B15, 5) PAC 687D8, 6) cosmid c62F6, M2) SPP1 marker. c62F6 was used as a positive control. The sizes and positions of the marker bands are indicated. DNA for all the clones was digested using restriction enzymes and the DNA fragments were resolved by PFGE within the range 5 kb - 150 kb. Subsequently, PFGE filter was prepared and probed with radiolabelled c62F6 probe. The restriction fragments that hybridised to the c62F6 probe are depicted as red dots. Out of the five clones screened three were real positives for the probe c62F6. **(C)** Sizes of clones as estimated from the restriction fragments that are indicated. Contig established based on Southern hybridisation results and shared restriction fragments is shown.

Figure 6.3 Fluorescent *in situ* hybridisation. **(A)** An example of a proximal FISH signal to the 16p13.3 breakpoint of t(14;16). The FISH signals obtained by the hybridisation of PAC 722G3 to the metaphase chromosomes of patient were seen on the normal chromosome 16 and the der(16) of the t(14;16) indicating that the PAC 722G3 is proximal to the 16p13.3 breakpoint of the t(14;16). **(B)** An example depicting a distal FISH signal to the 16p13.3 breakpoint obtained by the hybridisation of BAC 114I12 to the metaphase chromosomes of patient with t(14;16). The FISH signals were seen on the normal chromosome 16 and the der(14) indicating that the BAC 114I12 is distal to the 16p13.3 breakpoint of the t(14;16).


PAC722G3
N(16)
der(16)
a


BAC114I12
N(16)
der(16)
b
der(14)

6.3.2.3 FISH analysis of PAC clone 722G3

PAC DNA clone 722G3 was used for FISH analysis on the metaphase chromosomes of the patient with the t(14;16) *de novo* balanced translocation. The FISH signals for the PAC 722G3 were seen on both the normal chromosome 16 and the der(16) indicating that this clone was proximal to the 16p13.3 translocation breakpoint of t(14;16), Figure 6.3 A.

6.3.2.4 Mapping of probes in the vicinity of 16p13.3 breakpoint in CY182

Additional markers [D16S3088 (AFMb281ze9), A008S47 and WI-6421] were localized in the vicinity of the 16p13.3 breakpoint of the t(14;16) to generate more extensive physical map of the region. Information regarding these markers was obtained from the integrated map of human chromosome 16 available on the World Wide Web at URL W4, Table 1.1. These markers were mapped by PCR within the hybrid interval defined by CY182 and CY177, thereby separating and ordering the chromosome breakpoints in these two hybrids. The primer sequences for these markers are presented in Table 6.1.

An example of somatic cell hybrid analysis showing the localization of the DNA marker WI-6421 in the hybrid interval defined by CY182 and CY177 is presented in Figure 6.4. There were no markers identified earlier (Figures 3.1 and 4.7) separating these two breakpoints. The CY177 hybrid was derived by fusing the human cell line (TROJ) containing the der(16) of the familial translocation t(16;22)(p13;q11) with A9 mouse cell line (Callen et al., 1995). These results indicated that the order of the chromosome breakpoints of the hybrids at 16p13.3 was pter-23HA-CY196-CY197-CY182-CY177-CY198-CY168.

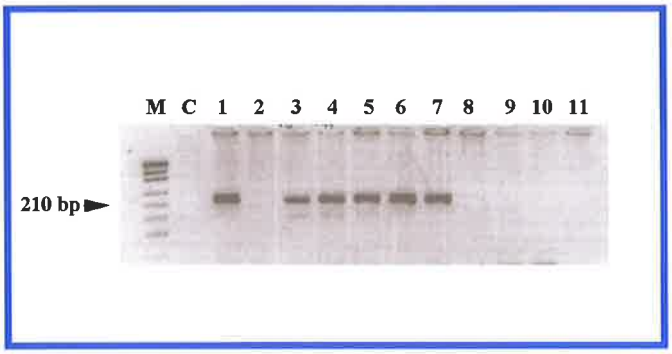


Figure 6.4 Mapping of WI-6421 with respect to the 16p13.3 breakpoint in CY182 using somatic cell hybrids at 16p13.3. Samples: 1) Human DNA, 2) A9, 3) CY18, 4) 23HA, 5) CY196, 6) CY197, 7) CY182, 8) CY177, 9) CY194, 10) CY198, 11) CY168. The CY18 and A9 somatic cell hybrids were used for determination of the chromosome 16 specific bands and mouse bands respectively. Lane C, represents no DNA (water) control and M is pUC19 marker DNA. WI-6421 was located on chromosome 16 in the hybrid interval defined by CY182 and CY177, thereby separating the chromosome breakpoints in these two hybrids. The size of PCR product is indicated.

Table 6.1 Primer sequences for the DNA markers located proximal to the 16p13.3 breakpoint in CY182 at 16p13.3.

DNA markers	Primer sequences (5'-3')		PCR product size (base pairs)
	Forward primer	*Reverse primer*	
c62F6	**cgt tcc ttc tca ggc att tgt c**	**gct caa gag tct gaa gtg gga g**	163
D16S3088	**ctc tga ata ggg tgg gga tg**	**aag gaa atc tgg ggt gta cg**	223
A008S47	**ata aga gta ttt tct ttc ccc ac**	**atc tca aaa gag gct gag aaa ga**	102
WI-6421	**acc aag aga tca gct gtc taa aca**	**ttc aga gcc ttc att tgg ct**	210
AC007222 (top)	**gcc ctt ttt gtc cat cag tc**	**ttt cag gga tgg gct aat tg**	198
AC007222 (bot)	**gga agg agg gtg tca ctt ca**	**cac ctt ggg aga ctc atc ct**	208
SP6 192K18	**aat gca tga agg tga tgc tg**	**gag ggt aag tcg gga cca at**	205
2603/bot	**gcc aac tcg act ctc tgc tt**	**gtg gag cca ctg ctc ctt ag**	202
2357P14 (T7)	**cat gca agc ttc cct gtt ct**	**agg ctg agt gag acc cat gt**	182
3052B24 SP6	**ggt tgc agc cca tgt aaa at**	**aag ctc cat ggt tgg tgt tc**	162
114I12 T7	**acc acc tct tgt gca gga at**	**gtg gtt tga caa tcg caa tg**	200
114I12 SP6	**gca ttt gct ggg tct tca tt**	**agg gtg gta gca tct cat gg**	192

'top' and 'bot' refer to the primers designed from the distal (top) and proximal (bottom) ends of the respective genomic sequences.

6.3.2.5 Identification of a cDNA clone and BAC DNA clone using DNA markers A008S47 and WI-6421

Having localized the DNA marker A008S47 in the hybrid interval defined by CY182 and CY177 further mapping information on this marker was obtained by accessing the GeneMap'99 data provided for this marker at http://www.ncbi.nlm.nih.gov/genemap/. The electronic PCR data provided for this marker indicated that A008S47 is present in the 5' end (accession number, H08623) of the Image human cDNA clone 44969. The 3' end sequence for this cDNA clone was also available in GenBank under the accession number, H08624, which was found to contain the DNA marker WI-6421. Information was also provided regarding a chromosome 16 BAC clone 441K8, whose working draft sequence was available in two ordered pieces (accession number, AC007222). The STS A008S47 was present within nucleotides 55807-55908 bp of the AC007222 sequence. The Image human cDNA clone 44969 and BAC DNA clone 441K8 were obtained from Genome Systems, St Louis, Missouri and Research Genetics, USA respectively for further analyses.

6.3.2.5.1 Image human cDNA clone 44969

Because the DNA markers A008S47 and WI-6421 were localized in the adjacent proximal interval of the 16p13.3 breakpoint contained in CY182 and were present at the 5' and 3'ends of the Image human cDNA clone 44969 respectively, the cDNA clone was obtained and processed to determine the corresponding transcript. The Image human cDNA clone 44969, which was derived from the Soares infant brain 1NIB cDNA library, is part of the human Unigene cluster number Hs.22833. However, the EST sequences present in the Unigene cluster corresponding to cDNA clones Image:44969, Image:34949 and c-0ca10 when

analyzed with respect to the genomic sequence (BAC clone 441K8; accession number, AC007222) had no indications of intron exon boundaries being present, poly-adenylation signal was absent and there was no similarity to any known proteins. On the basis of this information it was considered unlikely that these EST sequences were part of a real transcript. Furthermore, the insert of cDNA clone Image:44969 (isolated using the restriction enzymes *Not*I and *Hin*dIII) when hybridized to a Multiple Tissue Northern blot (Clontech) did not detect any corresponding transcript.

6.3.2.5.2 BAC DNA clone 441K8

The BAC DNA clone 441K8 (RPCI-11 Human BAC library) was used for FISH analysis on metaphase chromosomes of the patient with the t(14;16) to determine if this clone spans the 16p13.3 breakpoint of the translocation. However, the FISH analysis indicated that the clone 441K8 was proximal to the 16p13.3 breakpoint. Because the working draft sequence for this clone was available in the NCBI database in two ordered pieces (accession number, AC007222), this enabled the design of primers from the extreme ends of the AC007222 sequence. The primers (Table 6.1) were designed from the distal (top) and proximal (bottom) ends of the BAC 441K8 sequence. These ends were localized by PCR of the somatic hybrids. Both the ends of the BAC clone 441K8 were localized to the hybrid interval defined by CY182 and CY177, confirming the FISH localization.

6.3.2.6 Identification of DNA clones overlapping BAC clone 441K8

A database search was carried out with the available sequence (accession number, AC007222) of the BAC clone 441K8 for the identification of any overlapping sequenced clones. This BLAST search of the gss database (NCBI genomic survey

sequence database) identified the T7 end of BAC 3052B24 and TF end of BAC 2170H16, which were located at the extreme ends of the BAC 441K8 sequence. The sequence analysis indicated that these clones partially overlapped and extended beyond BAC 441K8. Therefore, BACs 3052B24 and 2170H16 (CIT_HSP, Human BAC library) were obtained through Research Genetics, USA, to perform FISH analysis on metaphase chromosomes of the patient with t(14;16) to determine if any of these clones spanned the 16p13.3 breakpoint. However, the FISH analysis indicated that the BAC 3052B24 was proximal to the 16p13.3 breakpoint. The clone BAC 2170H16 was located on chromosome 4q21.

6.3.2.7 Screening of the human PAC library (RPCI5)

DNA probes localized in the immediate vicinity of the 16p13.3 breakpoint of the t(14;16) were then used to screen a PAC library to identify additional clones to further extend the physical map. The distal probes selected to screen the PAC library were D16S3128 and T53018 (Figure 4.7), which were located outside the YAC 761C2 in the hybrid interval defined by CY197 and CY182. The proximal probes selected to screen the PAC library were D16S3088, A008S47 and WI-6421 and the end probes from BAC 441K8. All probes were first hybridized individually to filters of the PAC vector and human genomic DNA to ensure that cross hybridization was absent. The results indicated that only the probes T53018, WI-6421 and the proximal end probe generated from BAC 441K8 were suitable for hybridization to PAC filters as the other probes contained repeat sequences. Therefore, the probes T53018, WI-6421 and the proximal end probe generated from BAC 441K8 were used to screen the human PAC library RPCI5 (PAC/BAC Resources, Buffalo, New York). An example of the results obtained is depicted for RPCI5 filter number 20D containing a positive clone 953H7 in Figure 6.5 A.

Five PAC clones were identified, purchased, and DNA prepared. The clones were subsequently processed for Southern analysis to be confirmed as true positives. DNA was digested using restriction enzymes *Not*I and *Eag*I. The digested fragments were resolved using PFGE at the resolution limits ranging from 5 kb to 150 kb. The BAC clones 441K8 and 3052B24 were included to align these clones with the PAC clones. The restriction fragment patterns obtained are indicated in Figure 6.5 B. The clones were arranged in a possible contig on the basis of shared similar size restriction fragments. This contig was then confirmed by Southern hybridization analysis by hybridizing the PFGE filters with the probes used to screen the PAC library. The results indicated that out of the 5 clones identified from screening the RPCI5 PAC library only two clones, PAC 953H7 and PAC 1109N1, were positive for WI-6421. None of the five clones screened were positive for the other two probes, T53018 and the proximal end probe generated from BAC 441K8. These clones were therefore false positives generated from the original screening. The restriction fragments of the PACs 953H7 and 1109N1 and BACs 441K8 and 3052B24 that hybridized to WI-6421 are indicated in Figure 6.5 B. Arrangement of these clones in a contig on the basis of these results is shown in Figure 6.5 C. Because the PACs 953H7 and 1109N1 did not extended beyond the 441K8 and 3052B24 BAC contig (Figure 6.5 C) they were not used for subsequent FISH analysis.

In a further attempt to extend the established contig, additional database analysis was undertaken using the BAC 441K8 sequence.

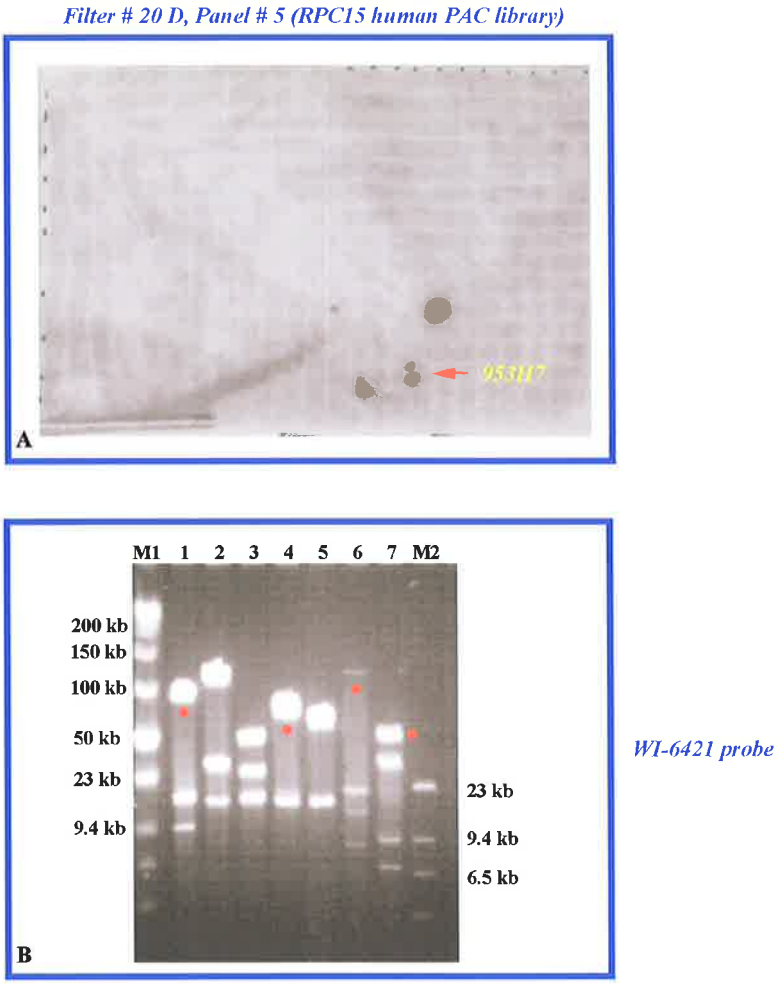


Figure 6.5 Screening of the human PAC library RPCI5 (PAC/BAC Resources, Buffalo, New York) with probes located proximal and distal to the 16p13.3 breakpoint of the t(14;16). **(A)** An example depicting a filter (number 20D) containing the positive clone 953H7, which is indicated by a red arrow. Only the panel (field section) for the respective filter containing the positive clone is shown. The filter number and panel number are indicated. **(B)** Restriction fragment analysis for PAC DNA clones screened using RPC15 PAC library. Samples: M1) low pulsed-field gel molecular weight marker, 1) PAC 953H7, 2) PAC 1127D2, 3) PAC 8274I22, 4) PAC 1109N1, 5) PAC 106B23, 6) BAC 441K8, 7) BAC 3052B24, M2) DNA marker Drigest. The sizes and positions of the marker bands are indicated. BACs 441K8 and 3052B24 were included for the purpose to align these clones with the PACs screened. DNA for all the clones was digested using restriction enzymes *Not*I and *Eag*I. Subsequently, the PFGE filters were prepared and hybridised to the probes used to screen the library. An example of the filter probed with WI-6421 is shown. The restriction fragments that hybridised to the probe are depicted as red dots. **(C)** The clones were arranged into a possible contig based on the Southern hybridisation results. The sizes of clones as estimated by PFGE are indicated.

6.3.2.8 Database searches using the BAC 441K8 sequence

As the Human Genome Project was rapidly generating genomic sequence regular BLAST searches of the human genomic sequence at NCBI was undertaken. One of these searches identified the BAC clone 192K18 (RPCI-11 Human BAC library). Complete sequence of this clone was available in GenBank under the accession number, AC006075. The SP6 and T7 end sequences of BAC 192K18 were also available under accession numbers, AQ417237 and AQ417240 respectively. Alignment of these (AC006075, AQ417237 and AQ417240) sequences with the BAC 441K8 sequence indicated that the SP6 end of the BAC 192K18 extended the contig previously generated using the BAC 441K8 sequence. At that stage the orientation of the contig was unknown. The order of these BAC clones in the contig on the basis of sequence alignment was BAC 3052B24-BAC 441K8-BAC 192K18.

Primers were designed (Table 6.1) from the SP6 end sequence of BAC 192K18 to map this end with respect to the 16p13.3 breakpoint by PCR using the somatic cell hybrids. The SP6 end of BAC 192K18 was still contained within the hybrid interval defined by CY182 and CY177, Figure 6.6.

A BLAST search of the gss database using the complete sequence of BAC 192K18 (accession number, AC006075) identified the BAC clone 3015A19 (CIT_HSP Human BAC library) whose SP6 and T7 end sequences were also available in GenBank under the accession numbers, AQ214317 and AQ091699 respectively. Alignment of the end sequences with BAC 192K18 (accession number, AC006075) sequence indicated that the SP6 end of BAC 3015A19 extended beyond the existing contig pter-(BAC 3052B24-BAC 441K8-BAC

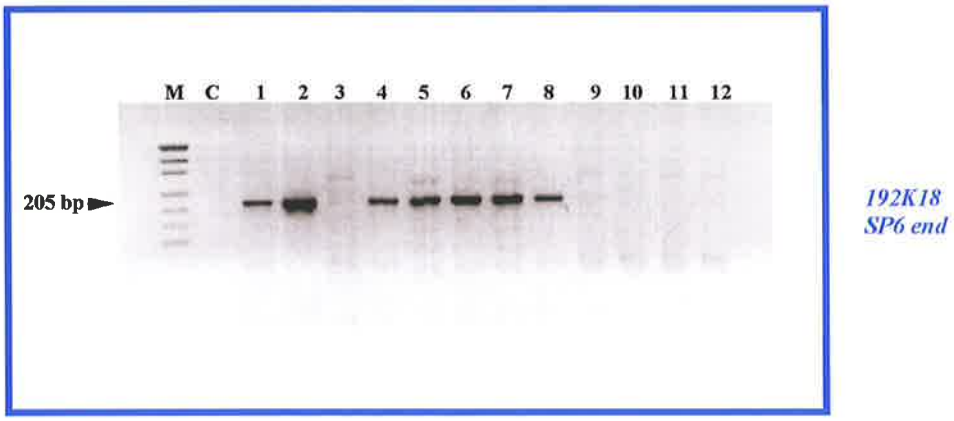


Figure 6.6 Mapping of SP6 end of BAC 192K18 with respect to the 16p13.3 breakpoint in CY182 using somatic cell hybrids at 16p13.3. Samples: 1) Human DNA, 2) BAC 192K18 DNA, 3) A9, 4) CY18, 5) 23HA, 6) CY196, 7) CY197, 8) CY182, 9) CY177, 10) CY194, 11) CY198, 12) CY168. The CY18 and A9 somatic cell hybrids were used for determination of the chromosome 16 specific bands and mouse bands respectively. Lane C, represents no DNA (water) control and M is pUC19 marker DNA. The BAC end was located in the hybrid interval defined by CY182 and CY177. The size of PCR product is indicated.

192K18- 3015A19). A further BLAST search using this SP6 end sequence of BAC 3015A19 identified an additional BAC clone 26O3 (RPCI-11 Human BAC library). Complete sequence for this clone was available in GenBank under the accession number, AC005774. Primers were designed (Table 6.1) from the end of this sequence, which did not overlap the existing contig, to map this clone with respect to the 16p13.3 breakpoint in CY182. PCR analysis of the somatic cell hybrids with this end probe of BAC 26O3 revealed that the clone was still contained within the hybrid interval defined by CY182 and CY177. With the identification of BAC clone 26O3 the order of the clones in the contig generated based on the sequence analysis data was pter-(BAC 3052B24-BAC 441K8-BAC BAC 192K18- 3015A19-BAC 26O3). The orientation of the contig was still not known at this stage.

Further database analysis with the BAC 26O3 sequence identified a BAC 2357P14 (CIT_HSP) whose SP6 end (accession number, AQ76666) was contained in the BAC 26O3 sequence. The low pass sequencing of BAC 2357P14 was available in 25 unordered contigs (accession number, AC027683), which was used for alignment of this clone with the existing contig. Additional searches of sequence database at NCBI with BAC 2357P14 did not extend the existing contig [pter-(BAC 3052B24-BAC 441K8- BAC 192K18- BAC 3015A19-BAC 26O3-BAC 2357P14)]. The alignment of these clones identified by BLAST search analyses with the sequence of BAC 441K8 was subsequently confirmed by restriction fragment and Southern analyses, Figure 6.10, with the identification of additional clones in the vicinity of the 16p13.3 breakpoint.

As expected from the PCR analysis, FISH analysis with the BAC DNA clones 192K18 and 2603 on the metaphase chromosomes of the patient with t(14;16) revealed that these BACs were proximal to the 16p13.3 breakpoint of the t(14;16).

6.3.2.9 Further extension of the 441K8 contig

Database searching with BAC 441K8 sequence allowed the contig shown in Figure 6.5 C to be extended as pter-(BAC 3052B24-BAC 441K8- BAC 192K18-BAC 3015A19-BAC 26O3-BAC 2357P14). The extremities of this contig were the T7 end (accession number, AQ076663) of the BAC 2357P14 and the SP6 end (accession number, AQ147393) of the BAC 3052B24. Primers were designed (Table 6.1) from either extremity and these both mapped by PCR to the hybrid interval defined by CY182 and CY177 (Figure 6.7 A and B). The SP6 end probe derived from BAC 3052B24 was also hybridized to the PFGE filter containing the BACs 3052B24 and 441K8 and confirmed that this end of BAC 3052B24 extends beyond BAC 441K8 (Figure 6.7 C).

The extending ends (SP6 end of the BAC 3052B24 and T7 end of the BAC 2357P14) of the contig were then used for another round of BAC library (RPCI-11 Segment 4) screening. The distal probe T53018 was also included in this screening because an earlier round of the PAC library RPCI5 screening (section 6.3.2.7) with probes located distal and proximal to the 16p13.3 breakpoint of the t(14;16) did not identify any clone positive for this (distal) probe. Identification of overlapping distal and proximal clones would help to determine the orientation of the contig across the 16p13.3 breakpoint.

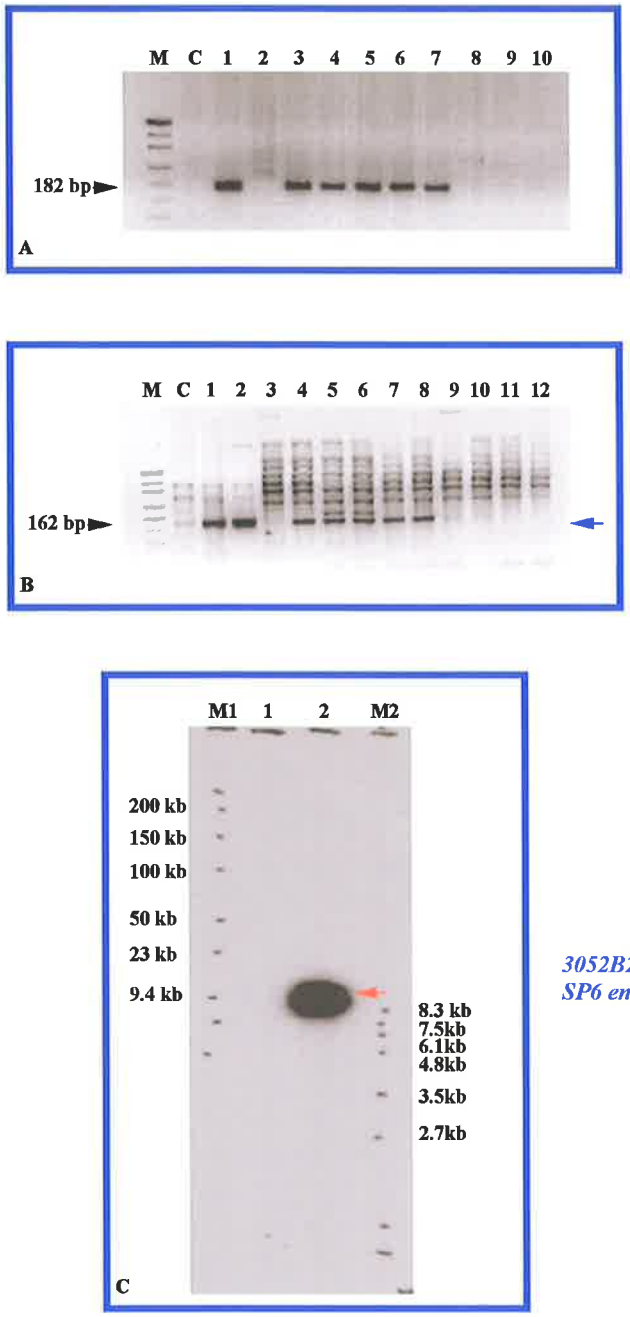
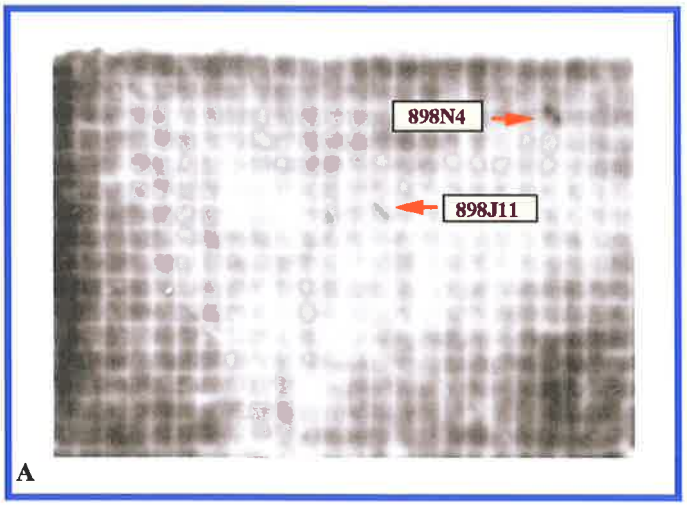


Figure 6.7 Mapping of the end probes of the contig generated from sequence analysis of BAC 441K8 sequence with respect to the 16p13.3 breakpoint in CY182 using somatic cell hybrids at 16p13.3. **(A)** Samples: 1) Human DNA, 2) A9, 3) CY18, 4) 23HA, 5) CY196, 6) CY197, 7) CY182, 8) CY177, 9) CY198, 10) CY168. **(B)** Samples: 1) Human DNA, 2) BAC 3052B24, 3) A9, 4) CY18, 5) 23HA, 6) CY196, 7) CY197, 8) CY182, 9) CY177, 10) CY194, 11) CY198, 12) CY168. The green arrow indicates the chromosome 16 specific bands. Both the extending ends of the contig were localised in the hybrid interval defined by CY182 and CY177. CY18 and A9 somatic cell hybrids were used for the determination of chromosome 16 specific bands and mouse bands respectively in both of the mapping experiments. In both the gels (shown in Figures A and B) lane C, represents no DNA (water) control and M is pUC19 marker DNA. The sizes of PCR products are indicated. **(C)** Mapping of SP6 end probe of BAC 3052B24 with respect to BAC 441K8. The end probe just hybridsed to itself (9.4 kb restriction fragment indicated by an arrow) which confirmed it to be the extending end on one end of the contig generated from sequence analysis of BAC 441K8.

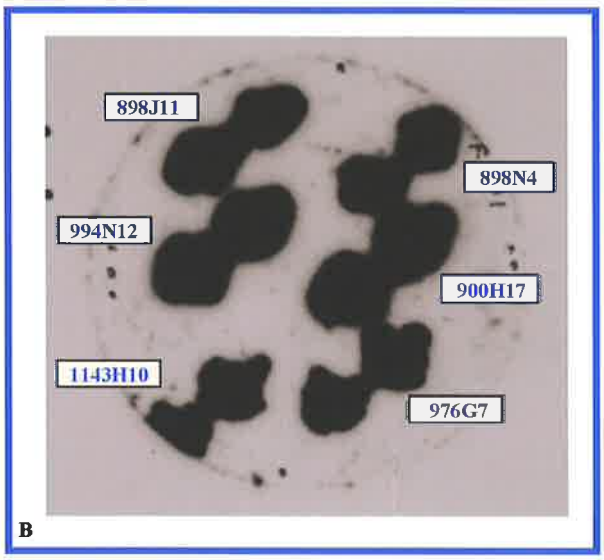
An example of the results with filter number 19K (RPCI-11 Segment 4 BAC library) containing positive clones is depicted in Figure 6.8 A. In all nine positive clones were identified and subsequently purchased (PAC/BAC Resources, Buffalo, New York). BAC 2357P14 was also purchased (Research Genetics) to be included as a positive control in the subsequent Southern hybridization analysis. These clones were initially confirmed as probable positives by colony blot hybridization using the same probes used to screen the BAC library. An example of the colony blot filter hybridized to α^{32}P-dCTP radiolabelled probe T53018 is shown in Figure 6.8 B. Subsequently, hybridization of Southern blots with α^{32}P-dCTP radiolabelled probes, T53018, SP6 end of the BAC 3052B24 and T7 end of the BAC 2357P14 showed that only four of these clones were true positives, 898J11, 9007H17, 994N12, and 1143H10 for the distal probe T53018 (Figure 6.10). None of these clones were true positives for the proximal probes (SP6 end of the BAC 3052B24 and T7 end of the BAC 2357P14). These clones were therefore false positives generated from the original screening. An example of a PFGE filter hybridized to the probe T53018 is shown in Figure 6.9 B.

DNA was prepared and digested using the restriction enzymes *Not*I and *Eag*I, and fragments resolved using PFGE at the resolution limits ranging from 5 kb to 150 kb. The BAC clones 3052B24, 441K8, 192K18, 26O3 and 2357P14 were included to align these clones with the BACs screened using RPCI-11 Segment 4 BAC library. An example of the restriction fragment patterns obtained is depicted in Figure 6.9 A. The clones were then arranged in a contig on the basis of shared similar size restriction fragments and probing with distal and proximal probes (Figure 6.10).

Filter # 19K, Panel # 4 (RPCI-11 segment # 4 human BAC library)



Probes:
SP6 3052B24
T7 2537P14
T53018



T53018

Figure 6.8 Screening of the human BAC library RPCI-11 segment # 4 (PAC/BAC Resources, Buffalo, New York) with probes: SP6 end of BAC 3052B24, T7 end of BAC 2357P14 and T53018. **(A)** An example depicting a filter (number 19K) containing the positive clones 898N4 and 898J11. Only the panel (field section) for the respective filter containing the positive clones is shown. The filter number and panel number are indicated. **(B)** Colony blot hybridisation. The positive BAC DNA clones were obtained and confirmed as probable positives by preparing colony blots. An example of the colony blot filter containing BAC clones screened from the BAC library RPCI-11 segment # 4 probed with T53018 is shown.

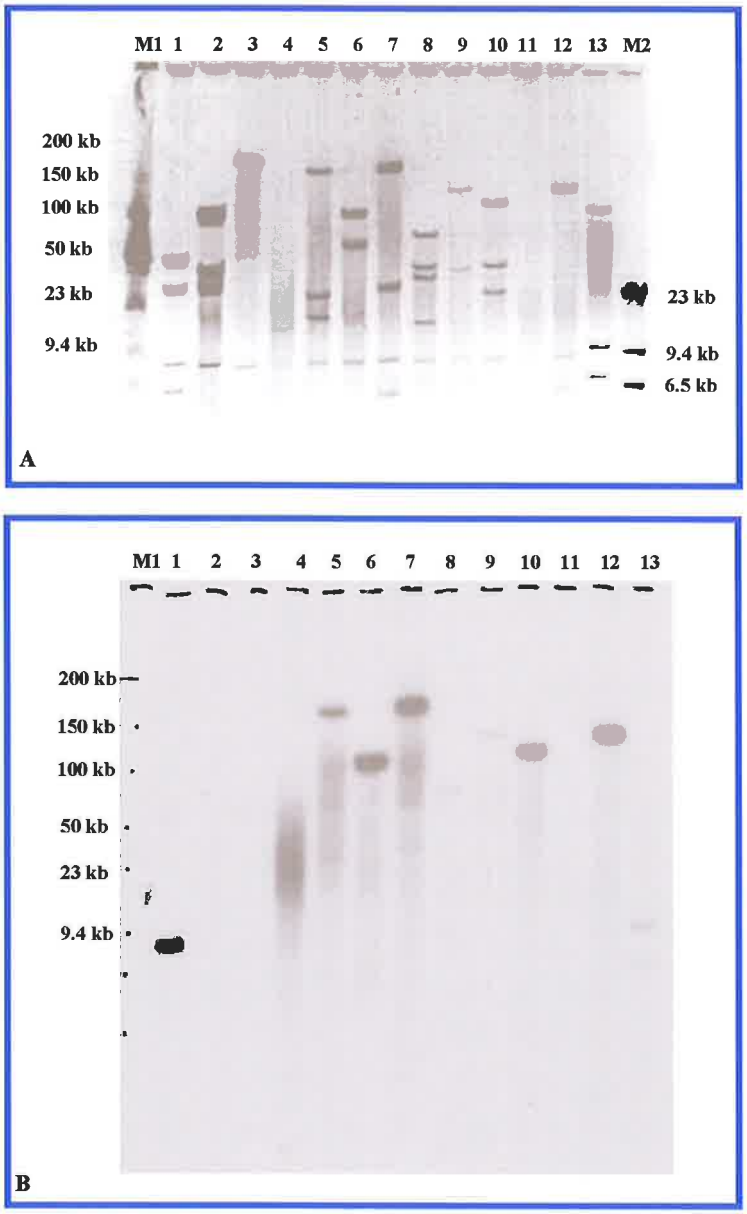


Figure 6.9 Restriction fragment and Southern analyses to arrange the clones screened from the human BAC library RPCI-11 segment 4 (PAC/BAC Resources, Buffalo, New York) into a possible contig. **(A)** Restriction fragment analysis. An example of the PFGE filter containing BAC clones from BAC library RPCI-11 segment 4 and those identified with the BLAST searches of BAC 441K8 sequence. Samples: M1) low pulsed-field gel molecular weight marker, 1) BAC 3052B24, 2) BAC 192K18, 3) BAC 778P19, 4) BAC 898N14, 5) BAC 898J11, 6) BAC 9007H17, 7) BAC 994N12, 8) 976G7, 9) 985E23, 10) 981H18, 11) 1032A14, 12) 1143H10, 13) 2357P14, M2) DNA marker lambda 4 *Hin*dIII. The sizes and positions of the marker bands are indicated. DNA for all the clones was digested using restriction enzymes *Not*I and *Eag*I. Subsequently, the DNA fragments were resolved by PFGE. The conditions for resolution of DNA fragments within the range 5 kb - 150 kb were: gel: 1% agarose gel in 0.5X TBE, running temperature: 14^{o}C, angles of electrodes: 120^{o}, time:15 hours. PFGE filters were prepared and hybridised with the probes used to screen the library. **(B)** An examlpe of the filter containing the DNA clones shown in Figure 6.9 A probed with T53018. Autoradiography was carried out at -70^{o}C for 4 hours. The clones were arranged into a possible contig based on the restriction fragment analysis and Southern hybridisation results of the filters hybridised with all three probes (SP6 end of BAC 3052B24, T7 end of BAC 2357P14 and T53018). The established contig and sizes of the clones as estimated by PFGE are depicted in Figure 6.10.

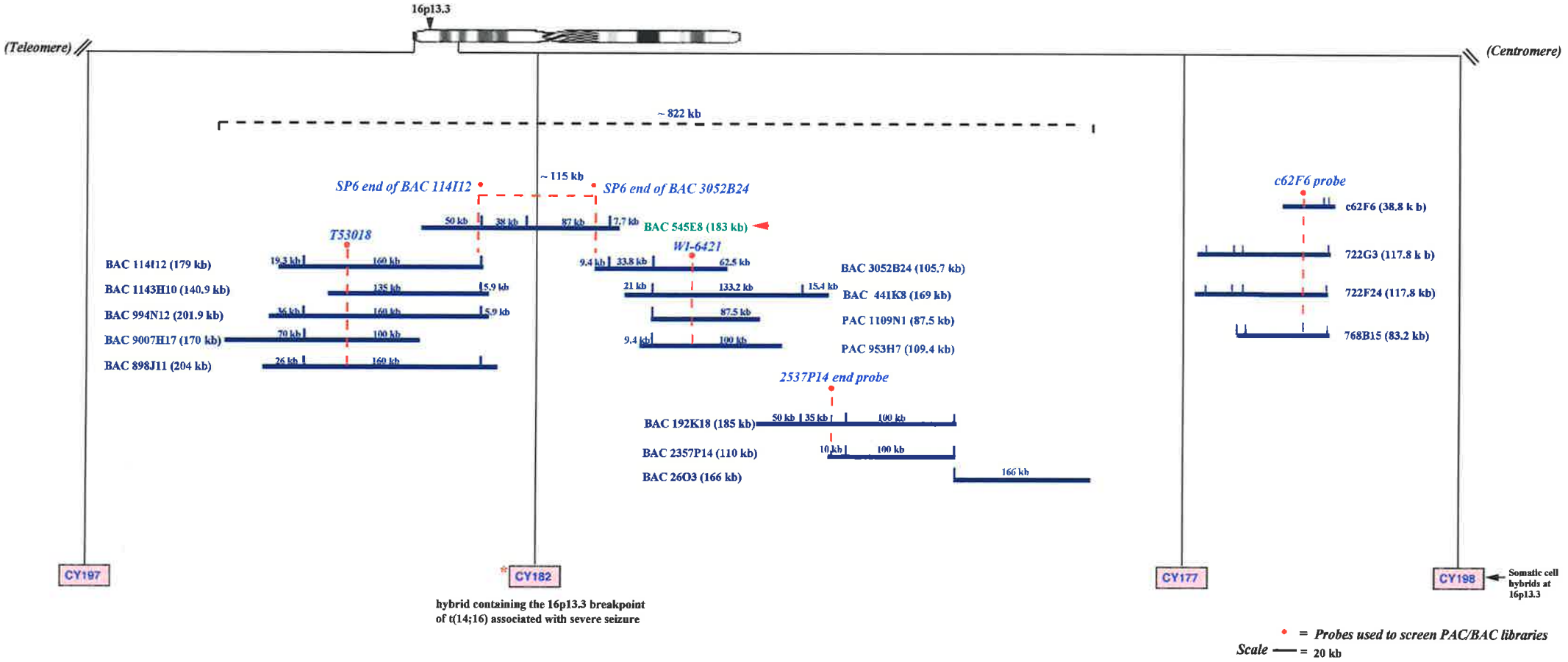


Figure 6.10 Positional cloning of the 16p13.3 breakpoint of the t(14;16). The BAC 545E8, which is indicated by an arrow was found to span the 16p13.3 breakpoint of the t(14;16) by FISH on metaphase chromosomes of the patient. The 16p13.3 breakpoint in CY182 was localised to the 115 kb distance on BAC 545E8 between the SP6 end probes of BACs 114I12 and 3052B24. The probes located distal to the 16p13.3 breakpoint in the hybrid interval defined by CY197 and CY182 and proximal to the 16p13.3 breakpoint in the hybrid interval defined by CY182 and CY177 were used to generate the PAC/BAC contig spanning the 16p13.3 breakpoint of the t(14;16). The position of the probes is depicted by red dots and dashed lines. The positions of chromosome breakpoints contained in the hybrids CY197, CY177 and CY198 with respect to CY182, which contains the der(16) of the t(14;16) characterised in the present study are also indicated. The contig spanning the 16p13.3 breakpoint of the t(14;16) spans over a genomic distance of 822 kb. The orientation of the contig across the 16p13.3 translocation breakpoint is from pter to centromere.

From the restriction map depicted in Figure 6.10 it was evident that the orientation of the BAC 2357P14 with respect to BACs 192K18 and 26O3 was opposite to that determined from the database searches using BAC 441K8 sequence (section 6.3.2.8). This was the result of the working draft sequence of BAC 2357P14 (accession number, AC027683) that was available as 25 unordered pieces. According to the Southern analysis the BAC 2357P14 was completely contained within the BACs 192K18 and 26O3 (Figure 6.10). Subsequently, this orientation of the BAC 2357P14 was further confirmed with the alignment of A2BP1 exons (Chapter 7) to the genomic clones identified in this region (Figure 7.4).

The RPCI-11 Segment 4 BAC library screening did not identify any BAC clone overlapping the distal and proximal contigs generated in the vicinity of the 16p13.3 breakpoint of the t(14;16). However, these contigs were later joined with the identification of BAC clones 114I12 and 545E8 (section, 6.3.2.10).

6.3.2.10 Identification of a clone spanning the 16p13.3 breakpoint of the t(14;16)

Subsequently, the NCBI database searches were also performed with the sequences for the DNA probe T53018 and the SP6 end of the BAC 3052B24. The marker T53018 identified an additional BAC 114I12 (RPCI-11 human BAC library). FISH analysis on metaphase chromosomes of the patient with t(14;16) revealed that the BAC 114I12 was distal to the 16p13.3 breakpoint, Figure 6.3 B. Because the SP6 (accession number, AQ346441) and T7 (accession number, AQ346442) end sequences were also available in GenBank, which enabled the design of primers (Table 6.1), this BAC was also localized by PCR analysis of the somatic cell hybrids at 16p13.3. Both ends were localized in the hybrid interval

defined by CY197 and CY182 similar to the marker T53018 (Figure 4.8), which identified this BAC.

The SP6 end of the BAC 3052B24 was the extreme end of the contig constructed proximal to the 16p13.3 breakpoint of the t(14;16). Database searches with this sequence identified the BAC clone 545E8 (RPCI-11 human BAC library). The working draft sequence for this clone was available as six unordered pieces in GenBank under the accession number AC022206. This sequence was analyzed to detect any matches with sequence located distal to the 16p13.3 breakpoint of the t(14;16). It was found that the sequence of the BAC 545E8 contained the SP6 end of BAC 114I12 (Figure 6.11 A). This finally was evidence that, the BAC 545E8 spans the breakpoint.

Southern analysis was performed to confirm these results. BAC 545E8 DNA was restricted with enzymes *Not*I and *Eag*I, analyzed by PFGE and probed with the SP6 end probes of the BACs 114I12 and 3052B24 (Figure 6.11 B). This Southern hybridization showed that the SP6 end probe of BAC 114I12 distal to the 16p13.3 breakpoint of the t(14;16) and the SP6 end probe of BAC 3052B24 proximal to the 16p13.3 breakpoint both hybridized to *Not*I and *Eag*I restriction fragments of BAC 545E8. Therefore, this confirmed that the BAC 545E8 crossed the 16p13.3 breakpoint of the t(14;16). From the restriction map, this breakpoint was localized within a distance of 115 kb on BAC 545E8 (Figure 6.10). The sizes and arrangement of BACs 114I12 and 545E8 are depicted in Figure 6.10. This allowed the contigs generated distal and proximal to the 16p13.3 breakpoint in CY182 to be orientated, Figure 6.10.

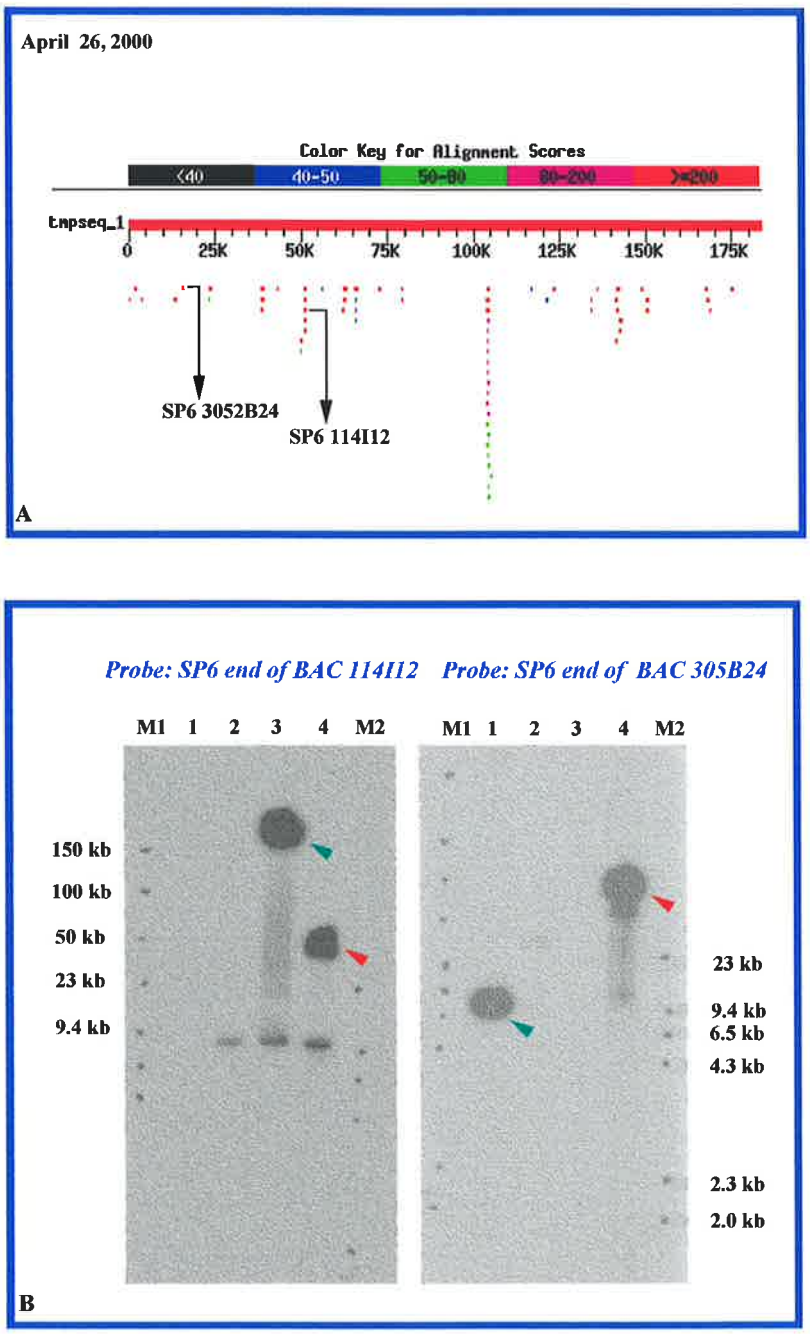


Figure 6.11 (A) BLAST search analysis of BAC 545E8 sequence through gss database at NCBI. The BAC sequence was found to contain the SP6 end of BAC 114I12, which was mapped distal to the 16p13.3 breakpoint of the t(14;16) and SP6 end of BAC 3052B24, which was mapped proximal to the 16p13.3 breakpoint of the t(14;16) indicating that the BAC 545E8 should span the 16p13.3 breakpoint of the t(14;16). The positions of SP6 end of BAC 3052B24 and SP6 end of BAC 114I12 are indicated. **(B)** Southern analysis was also done to confirm that BAC 545E8 contains the SP6 ends of BACs 3052B24 and 114I12. Samples: M1) low pulsed-field gel molecular weight marker, 1) BAC 3052B24, 2) BAC 192K18, 3) BAC 114I12, 4) 545E8, M2) DNA marker lambda 4 *Hin*dIII. The sizes and positions of the marker bands are indicated. The conditions for PFGE were: resolution limits: 5 kb - 150 kb, gel: 1% agarose gel in 0.5X TBE, running temperature: 14 ^{O}C, angles of electrodes: 120^{O}, time:15 hours. Green arrows indicate the hybridisation of the probe to itself and red arrows indicate the hybridisation of both the distal and proximal probes to BAC 545E8.

6.3.2.11 FISH analysis of BAC clone 545E8

To confirm these findings the BAC DNA clone 545E8 was used for the FISH analysis on the metaphase chromosomes of the patient with the t(14;16) *de novo* balanced translocation. The FISH signals for BAC 545E8 were seen on the normal chromosome 16, the derivative(16) and the derivative(14) (Figure 6.13 d) confirming that this clone spans the 16p13.3 translocation breakpoint.

6.3.2.12 Transcript analysis of the BAC 545E8 sequence

The BAC 545E8 sequence, which was available as six unordered pieces in GenBank under the accession number AC022206, was analyzed (June 15, 2000) using the BLAST algorithm for matches to ESTs (Figure 6.12). Most of these predicted ESTs were present as singletons, none of these ESTs had a poly A signal, the alignment of the predicted sequences available in the database was contiguous with the genomic sequence of BAC 545E8 and therefore, did not predict any exon/intron boundaries. On the basis of this information it was considered unlikely that the identified ESTs corresponded to a transcript. Therefore, expression analysis with these ESTs was not performed. There was one human Unigene cluster number Hs.165570 predicted on this BAC sequence. A human cDNA clone Image:25052 (Soares infant brain library) from this cluster was chosen for Northern analysis. However, this clone did not identify any bands. This cDNA was derived from an infant brain tissue and RNA from the brain tissue was contained on the Northern membrane, therefore it was considered unlikely that the transcript was expressed in tissues not present on the Northern. In addition the 5' (accession number, T81283) and 3' (accession number, R38891) sequences available in GenBank for this clone had no indication of intron exon boundaries being present, a poly-adenylation signal was absent and there was no

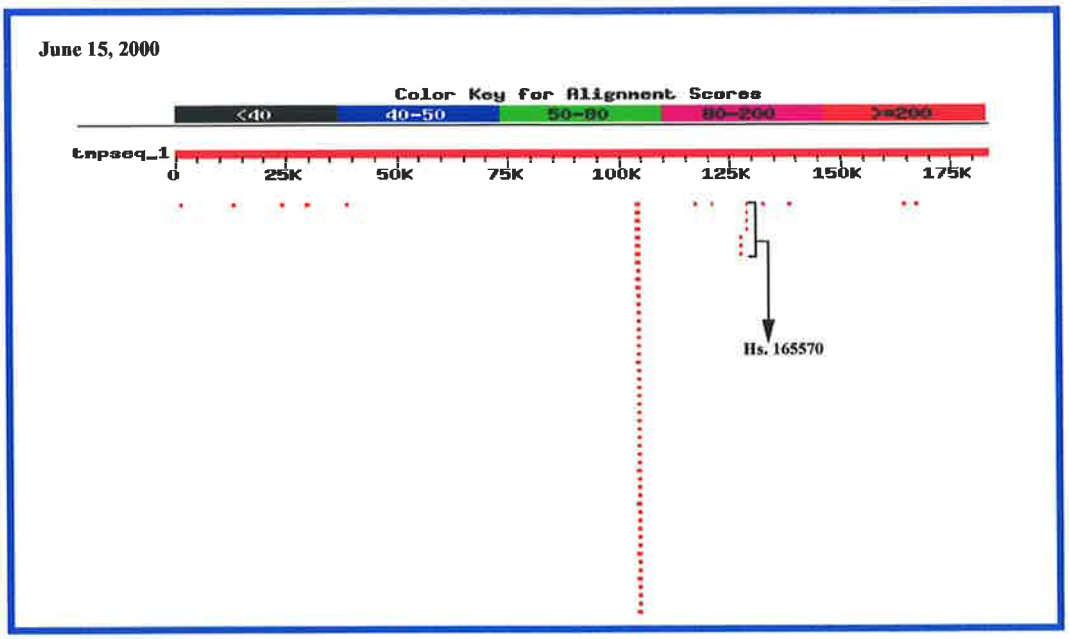


Figure 6.12 dbEST BLAST search analysis of BAC 545E8 sequence (accession number, AC022206) as on June 15, 2000. Most of the ESTs were predicted as singletons with only one Unigene cluster predicted on the sequence, which is indicated. cDNA clone from this Unigene cluster did not identify any transcript on Northern blot.

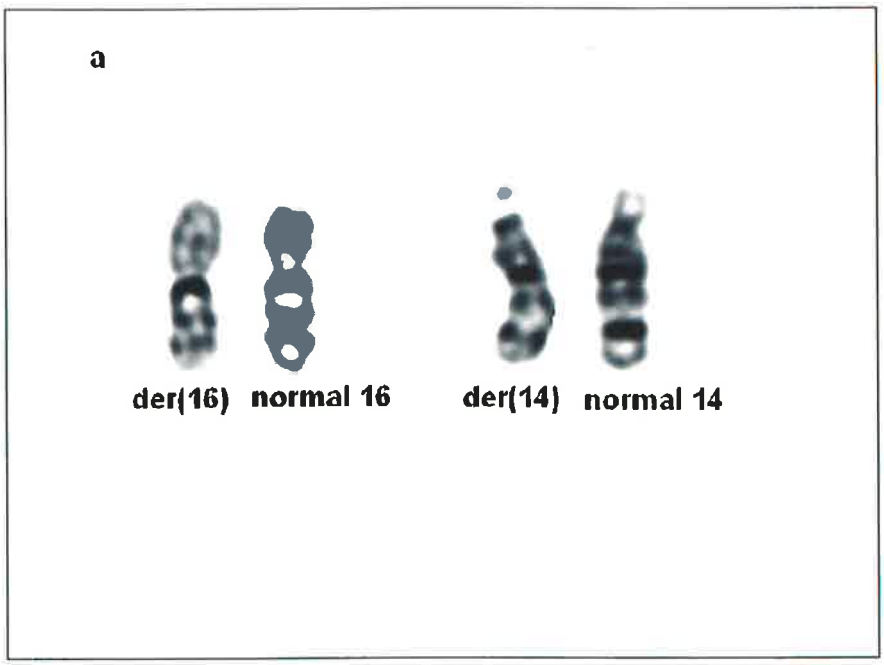
similarity to any known proteins. Together, these data reduced the possibility that this cDNA clone was a part of a real transcript. Later with the identification of A2BP1 gene (Chapter 7) it was revealed that both the chromosome 16p13.3 breakpoints of the t(1;16) and t(14;16) are located in introns of this gene.

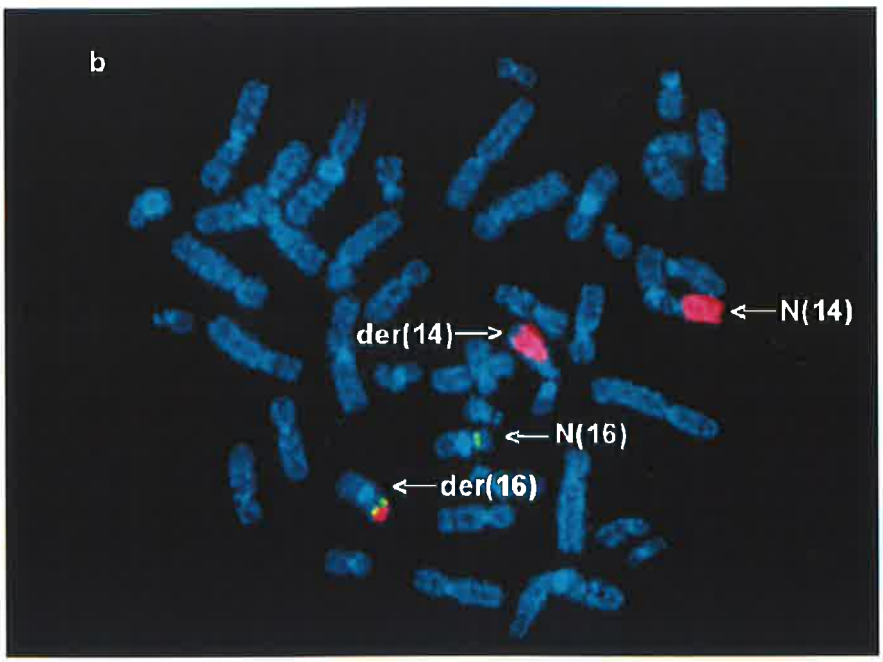
6.3.3 Resolution of chromosomal rearrangement in the patient initially detected with t(14;16)

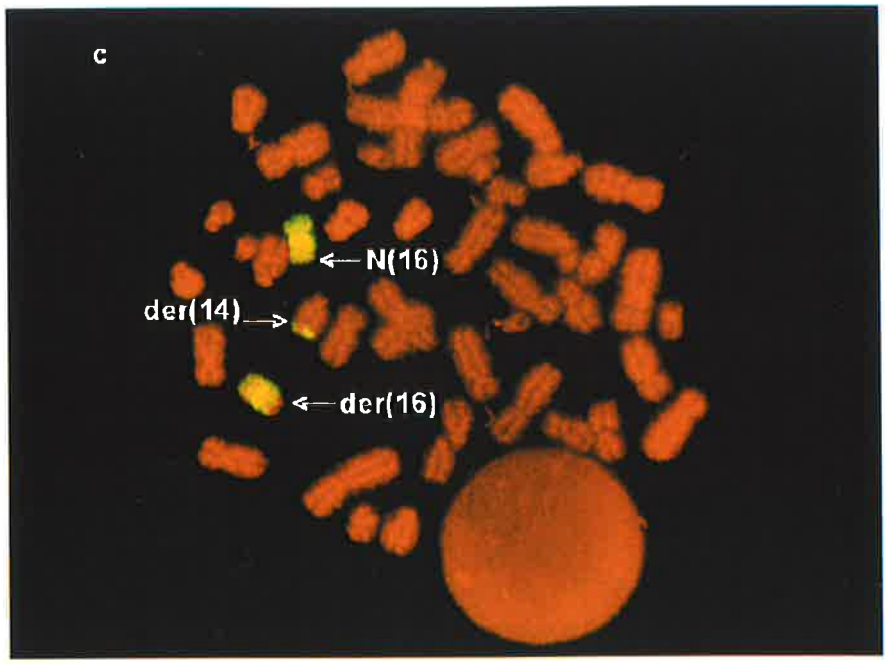
Classical G-banding from this patient demonstrated a *de novo* translocation between chromosomes 14 and 16, t(14;16)(q32;p13.3), Figure 6.13 a. There was no evidence from PCR localization of probes of any loss of material on chromosome 16 proximal to the 16p13.3 breakpoint in CY182. Chromosome painting with chromosome 14 and 16 paints (Figure 6.13 b and c) supported the interpretation of an apparent reciprocal translocation.

However, DNA clones which were localized proximal to the 16p13.3 breakpoint of the t(14;16) by FISH analysis on the metaphase chromosomes of the patient indicated that the translocation in this patient was also associated with an inversion of chromosome 16. The BAC clone 545E8, which was known to span the 16p13.3 breakpoint resulted in a single FISH signal on the normal chromosome 16 and as expected two signals on the translocated chromosome 16, one on the der(14) and the other on the der(16). However, unexpectedly the signal on the der(16) was at 16p11.2 rather than at 16p13.3. This suggested that the short arm of chromosome 16 involved in the translocation was also involved in a paracentric inversion with the breakpoints of the two arrangements at 16p13.3 being coincident.

Figure 6.13 FISH studies of chromosome 16 rearrangement. **(a)** GTL-banding of the normal and derivative chromosomes 14 and 16. **(b)** Cohybridisation of number 14 paint and b268E9. Number 14 paint was detected with TRITC (red). b268E9 was detected with FITC (green). **(c)** Number 16 paint detected with FITC (yellow). **(d)** Hybridisation signal from BAC 545E8 on the normal chromosome 16, the derivative 16 and the derivative 14 (spanning the translocation breakpoint).


a
der(16)
normal 16
der(14)
normal 14


b
der(14)
N(14)
N(16)
der(16)


c
N(16)
der(14)
der(16)

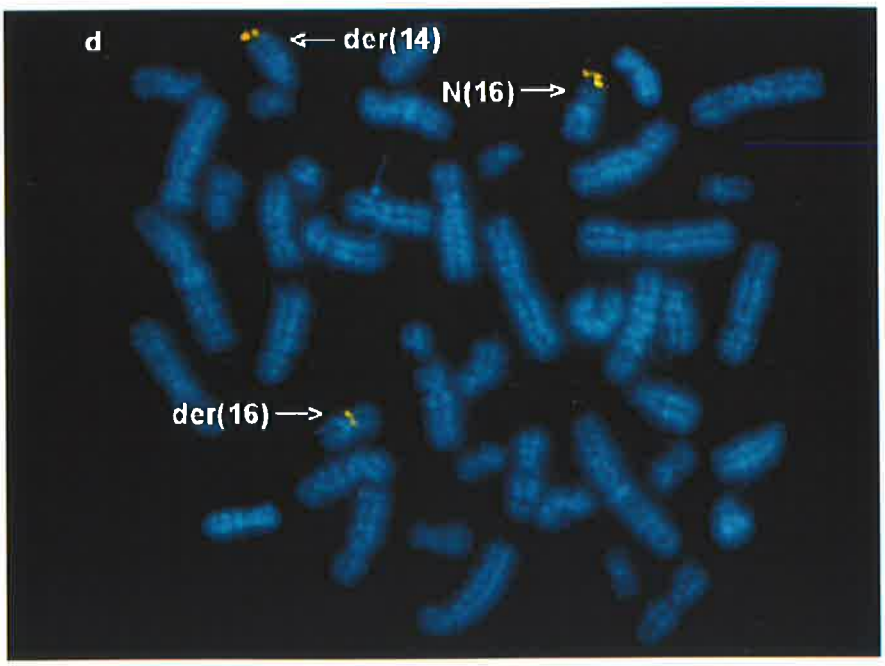

d
der(14)
N(16)
der(16)

Table 6.2 Resolution of der(16) by FISH

Probe	Somatic cell hybrid interval	Chromosome 16 band	FISH signal[1]
c77E8	CY14-CY190	p13.3	*Distal*
p102J11	23HA-CY196	p13.3	*Distal*
b375G12	23HA-CY196	p13.3	*Distal*
b315L9	CY196-CY182	p13.3	*Distal*
b118C2	CY196-CY182	p13.3	*Distal*
b114I12	CY197-CY182	p13.3	*Distal*
b545E8	CY197-CY182-CY177	p13.3	SPANS
b3052B24	CY182-CY177	p13.3	*Proximal,**
b441K8	CY182-CY177	p13.3	*Proximal,**
b192K18	CY182-CY177	p13.3	*Proximal,**
b26O3	CY182-CY177	p13.3	*Proximal,**
c62F6	CY177-CY198	p13.3	*Proximal,**
p722G3	CY177-CY198	p13.3	*Proximal,**
c10B8	CY177-CY198	p13.3	*Proximal,**
c37C6	FRA16A-CY183	p13.11	*Proximal,**
c307G2	CY13-CY15	p12.3	*Proximal,**
bA-279B10	CY156-CY165	p12.2	*Proximal,**
bA-268E9	CY156-CY165	p12.2	*Proximal,**
c311D5	CY155-CY160(D)	p12.1	*Proximal,**
bA-218C7	CY155-CY160(D)	p12.1	*Proximal,**
bA-485G10	CY155-CY160(D)	p12.1	*Proximal,**
bA-670B5	CY160 (D)-FRA16E	p12.1	*Proximal,**
b2049O4	FRA16E-CY12	p11.2p12.1	*Proximal*
bA-331G1	FRA16E-CY12	p11.2p12.1	*Proximal*
b761H5	FRA16E-CY12	p11.2p12.1	*Proximal*
bA-305A8	FRA16E-CY12	p11.2p12.1	*Proximal*

[1]Position of signal on short arm of chromosome 16 relative to translocation breakpoint at 16p13.3; (b) represents BACs and (p) PACs.

The order of probes is given from 16pter to centromere. The somatic cell hybrids are as given in Callen et al., (1995). The FISH signal designated as distal was located on the der(14), while proximal was located on the der(16). Those signals marked with an asterisk were located in a different position relative to the centromere on the der(16) compared with the normal chromosome 16 in the same metaphase.

This was confirmed by FISH using a series of DNA clones (Table 6.2) that were identified proximal and distal to the 16p13.3 breakpoint of t(14;16), probes known to be distributed on the short arm of chromosome 16 from their location on the mouse/human somatic hybrid map (Callen et al., 1995) and BACs from the BAC map of Cao et al., (1999). Six probes at 16p13.3 distal to the BAC 545E8 were only present on the der(14). A series of 15 probes between 16p13.3 and 16p11.2 were all located on the der(16) in positions relative to the centromere which were different from that on the normal 16p. These locations were consistent with the presence of a paracentric inversion. Four BACs at 16p11.2-12.1 were located at the same position on the normal and the der(16). The proximal breakpoint of the inversion was located between the BACs A-670B5 and 204904.

These data are consistent with a paracentric inversion of a portion of the short arm of 16 with the distal breakpoint coincident with that of the translocation at 16p13.3 and the proximal breakpoint at 16p12.1. This complex three break rearrangement can be notated as t(14;16)(p32;p13.3)inv16(p13.3p12.1).

6.4 Discussion

The t(14;16) *de novo* balanced translocation associated with mild mental retardation and grand mal seizures was characterised using a positional cloning strategy. This case provided a unique resource to identify gene(s) at or near the chromosomal breakpoint, the functions of which may be disrupted by the rearrangement and which may be responsible for the disease phenotype observed (Bedell et al., 1996; Baere et al., 2000). Studies involving patients with translocations associated with a particular disease has earlier led to the cloning of numerous disease causing genes (reviewed by Collins 1992, 1995).

In order to identify the disease gene responsible for the observed phenotype in the patient with t(14;16), the cosmid c62F6, which was located in the hybrid interval defined by CY182, CY177 and CY198 was initially used to screen the human PAC library RPC14 to construct a BAC/PAC contig across the 16p13.3 translocation breakpoint. The PAC clone 722G3 from this library was used to do FISH analysis on the metaphase chromosomes of the patient with t(14;16) which revealed that this clone was proximal to the 16p13.3 breakpoint of the t(14;16).

Additional markers [D16S3088 (AFMb281ze9), A008S47 and WI-6421] were then identified to be used as probes for screening of PAC/BAC libraries. These markers were localized within the hybrid interval defined by CY182 and CY177, thereby separating the chromosome breakpoints in these two hybrids. There were no markers identified earlier separating these two breakpoints in the previously reported map of chromosome 16 (Callen et al., 1995). The order of the breakpoints of the hybrids was refined to pter-23HA-CY196-CY197-CY182-CY177-CY198, CY168.

The STS A008S47 identified a BAC DNA clone 441K8. Analysis with the sequence for this clone identified another extending BAC clone 3052B24. However, both these clones were also proximal to the 16p13.3 breakpoint of the t(14;16). Therefore, the markers mapping distal and proximal to the 16p13.3 breakpoint were used for a second round of screening of a PAC library (RPCI5). The PAC clones 953H7 and 1109N1 were identified, which were contained within the BACs 441K8 and 3052B24 and were therefore not used for FISH analysis on the metaphase chromosomes of the patient.

Further analysis with the available sequence (accession number, AC007222) of BAC clone 441K8 identified additional overlapping DNA clones. These BAC clones were arranged in a contig pter-(BAC 3052B24-BAC 441K8- BAC 192K18- BAC 3015A19-BAC 26O3-BAC 2357P14) based on the sequence analysis data. The BAC clones 192K18 and 26O3 identified by these searches were subjected to FISH analysis on the metaphase chromosomes of the patient. However, both the clones were proximal to the 16p13.3 breakpoint. This localization was also confirmed by PCR analysis of the somatic cell hybrids at 16p13.3. All the clones were found to map in the hybrid interval defined by CY182 and CY177.

Therefore, a third round of the BAC/PAC library screening was undertaken. The BAC library RPCI-11 Segment 4 was screened using the probes proximal to the 16p13.3 breakpoint, the end probes (T7 end of the BAC 2357P14 and SP6 end of the BAC 3052B24) of the contig generated using the database searches with the BAC 441K8 sequence. In addition, the probe T53018, which was localized distal to the 16p13.3 breakpoint of the t(14;16) was also used. The positive clones screened were arranged in the contig (Figure 6.10) on the basis of the restriction fragment and Southern analyses. However, this screening did not identify any clone spanning the 16p13.3 breakpoint.

As the HGP gained momentum there was increasing availability of human genome sequence in the NCBI database. Database searches using the sequences of the SP6 end of the BAC 3052B24 and the T53018 probe eventually identified the BAC clones 114I12 and 545E8. The BAC 545E8 was found to span the 16p13.3 breakpoint contained in CY182 by FISH analysis on the metaphase chromosomes

of the patient with the t(14;16). The DNA clones identified distal and proximal to this 16p13.3 breakpoint were arranged in a contig that is depicted in Figure 6.10, which spanned over approximately 822 kb of genomic DNA.

Analysis of the BAC 545E8 sequence identified a number of matching ESTs. However, these ESTs were present as singletons, did not contain a poly A signal, and were contiguous with the genomic sequence. Therefore, it was considered unlikely that these ESTs were part of a real transcript. Two cDNA clones, 44969 and 25052, identified with sequences of BACs 441K8 and 545E8 respectively, did not identify any transcript when hybridized to a Northern blot. With the identification of the A2BP1 gene (Shibata et al., 2000) it was revealed that the sequenced regions of the BACs 441K8 and 545E8 that were used for expression analyses reside in introns of the A2BP1 gene (Chapter 7).

The t(14;16)(q32;p13.3) was characterized in detail using FISH with a BAC 545E8, which was found to span the 16p13.3 breakpoint and a series of probes located proximal and distal to this breakpoint. These results demonstrated a complex three break chromosomal rearrangement, which was apparently derived from a simultaneous translocation and paracentric inversion. This was suggested by the sharing of a breakpoint between the two rearrangements. The chromosomal abnormality is notated as t(14;16)(p32;p13.3)inv16(p13.3p12.1). Translocations and inversions of chromosome 16 have been previously described in the patients with M4Eo acute myelomonocytic leukaemia (Maarek et al., 1999). This unexpected complex three break chromosomal rearrangement in the patient initially reported with the t(14;16) complicated the further positional cloning approach.

Expression and Mutation analyses of C16*orf*5 Gene and Identification of A2BP1 disrupted by the t(1;16) and t(14;16)

Chapter 7

7.1 Introduction

A patient was reported at the Department of Cytogenetics, WCH, Adelaide, with a balanced *de novo* translocation 46,XY,t(1;16)(q12;p13.3) associated with severe mental retardation. The 16p13.3 breakpoint of this translocation was characterized (Chapter 5) which failed to identify any transcript at the breakpoint. However, there are cases where translocation breakpoints map outside the putative disease genes (Kleinjan and Heyningen, 1998). Chromosomal rearrangements frequently lead to alteration of the gene's environment and this may be reflected in a change in the expected level of expression, referred to as a position effect (Milot et al., 1996). The chromosome 1 breakpoint of the t(1;16) was shown to be within the pericentromeric heterochromatin (section, 5.3.2). It was possible that the transcription of the gene C16*orf*5 (Chapter 3), which is the closest identified gene distal to the 16p13.3 breakpoint of the t(1;16), was altered due to the position effect of the adjacent translocated chromosome 1 heterochromatin. Examples demonstrating the silencing or altered expression of a gene and its proximity to a heterochromatic compartment are reviewed in Milot et al., (1996) and Cockell and Gasser, (1999).

Because C16*orf*5 is the closest gene to the breakpoint, and has an abundant brain specific expression, this was considered an excellent candidate for involvement in non specific-MR. The hypothesis is that there is a position effect of the adjacent translocated chromosome 1 heterochromatin on the expression of the C16*orf*5 gene. Therefore, the expression of the C16*orf*5 gene was compared between normal diploid tissue and tissue with translocation by semi-quantitative RT-PCR analysis (Bercovich et al., 1999). Further, to demonstrate any possible role of

C16*orf*5 gene in mental retardation, the open reading frame of this gene was screened for mutations in 100 female patients with non-specific MR.

Subsequent to this work being completed a database search performed on November 11, 2000 using the BAC 375G12 DNA sequence revealed that the 5' UTR of the novel protein A2BP1 was contained in the BAC 375G12. This protein was recently isolated by yeast two-hybrid screening using an ataxin-2 C-terminal as "bait" (PCR amplified 3' cDNA fragment corresponding to a C-terminal fragment of ataxin-2 subcloned into pBDGal4Cam vector) by Pulst and colleagues (2000). Ataxin-2 is the gene product of human SCA2 gene that is known to cause Spinocerebellar ataxia type 2 (Huynh et al., 1999). A2BP1 is predominantly expressed in brain and muscle and binds to the C-terminus of ataxin-2 (Shibata et al., 2000). Co-localization of A2BP1 and ataxin-2 to the trans-Golgi network and dentate neurons suggest that the specific binding of A2BP1 to ataxin-2 in brain may contribute to the restricted pathology of SCA2. The patients with SCA2 often complaints of muscle cramps, which may be related to ataxin-2 interaction with A2BP1 (Shibata et al., 2000). The physical mapping data of the breakpoints generated in the present study together with the A2BP1 cDNA sequence suggests that both the chromosome 16p13.3 breakpoints of the t(1;16) and t(14;16) are located in introns of A2BP1 gene.

This Chapter presents the expression and mutation studies of the C16*orf*5 gene and the physical mapping data of A2BP1 with respect to the 16p13.3 breakpoints of the t(1;16) and t(14;16).

7.2 Methods

The methods specific for this Chapter are presented. Other general techniques used in this study are presented in Chapter 2.

7.2.1 Isolation of total RNA

An established lymphoblastoid cell line (MI-APY) in the Department of Cytogenetics and Molecular Genetics, WCH, Adelaide from the patient with t(1;16) was used for the isolation of total RNA to perform semi-quantitative RT-PCR analysis of C16*orf*5 gene in this patient. RNA was also isolated from another established lymphoblastoid cell line (SKB) from an individual with a normal karyotype to be used as control in RT-PCR experiments. RNA was isolated by using either Trizol reagent method or RNeasy Midi Kit (Qiagen) columns. Both the procedures were carried out under RNAase free conditions. The solutions used were incubated in DEPC (0.2%, v/v) with constant shaking at room temperature. Subsequently the solutions were autoclaved. Filtered Gilson pipette tips were used and the Gilson pipettes were cleaned with RNaseZAP (Ambion) prior to use in isolation of RNA. Gloves were worn throughout the procedure. Isolated RNA was stored at –70°C.

7.2.1.1 Total RNA isolation using Trizol reagent

Total RNA using Trizol reagent was isolated using a method based on that of Chomczynski and Sacchi (1987). $1X10^7$ cells obtained from cell cultures were used as a starting material for total RNA isolation. The cells were washed in PBS and spun down at 1,200 rpm for 10 minutes. The cells were then resuspended in 1 ml of Trizol by brief vortexing. One-tenth of the volume of chloroform was added. Sample was vortexed briefly for 25 seconds and then placed on ice for 10

minutes. The sample was then spun at 12,000 for 15 minutes at 4°C. The aqueous phase was then collected in a fresh tube. To this, an equal volume of isopropanol was added and the sample was incubated on ice for 15 minutes followed by a centrifugation at 4°C for 15 minutes. The supernatant was discarded and the pellet was washed by adding at least 1 ml of 75% ethanol per 0.75 ml of Trizol reagent used for the initial homogenization. The contents in the tube were mixed by gently tapping the tube with finger and centrifuged at 7,500 rpm for 2-5 minutes at 4°C. The supernatant was removed and the RNA pellet was briefly air-dried for 10-15 minutes. The RNA pellet was resuspended in 50 µl of DEPC-treated water by passing the suspension a few times through the pipette tip to ensure complete resuspension of the RNA pellet.

The resuspended RNA was reprecipitated by adding one-twentieth the volume of 4 M NaCl and 2 volumes of 100% ethanol. This second precipitation was recommended for RNA to be used in PCR and enzymatic reactions. The reaction was incubated at –20°C for one hour followed by a centrifugation at 7,500 rpm for 15 minutes at 4°C. The RNA pellet was washed again with 75% ethanol, air-dried for 20 minutes and resuspended in DEPC treated-water. This was then followed by the quantitation of diluted (1:100 in DEPC water) RNA using spectrophotometry (section, 2.3.1.6). The concentration of RNA sample was determined by measuring the absorbance at 260nm (A_{260}). In case of RNA the relevant conversion factor is 40, which gives a concentration in µg/ml. The RNA concentration was calculated using the following formulae.

RNA yield in µg/ml = OD reading X Dilution factor (100) X Conversion factor (40) X (Total volume of RNA sample prepared)/1000.

An aliquot of the purified RNA (3 to 4 µl) prepared was then subjected to agarose gel (0.8%) electrophoresis in order to determine its integrity and size distribution. Isolated RNA was stored at –70°C.

7.2.1.2 Total RNA isolation using RNeasy Midi Kit (Qiagen)

Total RNA using RNeasy Midi Kit was performed as per the protocol described in RNeasy Midi handbook catalogue number, 75142. The amount of starting material recommended was 3-4X10^7 cells for high and pure RNA yield with RNeasy Midi Qiagen columns (Qiagen, Hilden, Germany). Cells obtained from cell cultures were washed in PBS and spun down at 1,200 rpm for 10 minutes. The cell pellet was loosen by gently flicking the tube with finger. This was done to allow complete cell lysis for maximum RNA yield. For cell numbers less than 3X10^7, 1.9 ml of kit supplied RLT buffer (10 µl of 14.5 M β-ME was added per 1 ml of buffer RLT prior to use) was added to the cells. The sample was then briefly vortexed (10 seconds) and homogenized by passing it 5-10 times through first a 21 gauge needle and then a 18 gauge needle fitted to an RNase free syringe. This allowed complete homogenization of the cell lysate. Incomplete homogenization will lead to significant reduced yields and can cause clogging of the RNeasy midi spin column.

To the homogenized lysate was added 3.8 ml (approximately one volume of the lysate) of 70% ethanol. The sample was then mixed by brief vortexing. The sample was then applied to RNeasy midi spin column fitted with a 15 ml centrifuge tube (kit supplied) and the tube was closed gently. This was followed by centrifugation for 10 minutes at 3,000 rpm, to allow the lysate to completely pass through the column. The flow-through was discarded and the centrifugation

tube was reused in successive steps till the sample was ready for the RNA elution step. To the column was then added 3.8 ml of buffer RW1 (kit supplied), the centrifugation tube was closed gently and spun at 5 minutes to wash the column. The flow-through was discarded again and 2.5 ml of the buffer RPE (kit supplied) was added to the column and the column was spun for 2 minutes at 3,000 rpm (prior to the use of buffer RPE 4 volumes of 100% ethanol was added to obtain the working concentration). This step was repeated again to dry the spin column membrane because any traces of residual ethanol may interfere with subsequent reactions.

The RNeasy midi spin column was then transferred to a new 15 ml collection tube (kit supplied) and total RNA was eluted by adding 150 µl of (kit supplied) RNase-free water. The column was incubated for 1 minute at room temperature and then was spun for 3 minutes to elute total RNA. The elution step was repeated for maximum RNA recovery. The concentration, integrity and size distribution of the total RNA sample prepared was determined as described in section 7.2.1.1. Isolated RNA was stored at –70°C.

7.2.2 RT-PCR analysis

RT-PCR analysis was performed as described in section 2.3.8 to determine the expression of C16*orf*5 gene in lymphoblastoid cells using the gene specific (GSP) primers. The primer pair GSP#1 was derived from WI-16589 EST (Chapter 3, section 3.3.2). The forward and reverse sequences for WI-16589 primer pair are provided in Table 3.1. The forward and reverse sequences for primer pair GSP#2 are as follows: forward primer 5' GGT CTC ACA CCC TCA GTG CT 3' and

reverse primer 5' ACT GAC CAA GAC TGC CAC CT 3'. These primers amplified a 430 bp product.

7.2.3 Semi quantitative RT-PCR analysis

RNA was prepared as described in section 7.2.1. 1 μg of the total RNA was reverse transcribed for first-strand cDNA synthesis followed by standard PCR using reagents from Gibco-BRL (Life Technologies, Gaithersburg, USA, see section, 2.3.8). Standard PCR using 1/10 of the first strand cDNA was performed in a final volume of 20 μl with 0.5 units of Taq DNA polymerase (Gibco-BRL, Life Technologies, Gaithersburg, USA). Initially the PCR was performed at 20, 25, 30 and 35 numbers of PCR cycles using the identical quantities of the patient and control (normal lymphoblastoid cell line, SKB, section 7.2.1) template DNA and 150 ng/μl of the control (esterase D) and GSP primer pairs. This was done to determine the minimum number of PCR cycles needed to visualize the RT-PCR products on agarose gel. A minimum of thirty PCR cycles were initially chosen for subsequent RT-PCR experiments. For quantitative PCR amplification reactions, internal standards, in which the sample and the internal control are amplified in the same tube (duplex amplification reaction, DA), are preferred (Bercovich et al., 1999). The internal standard approach eliminates the differences in the reaction conditions and ensures reliable internal controls. Therefore, having standardized the number of PCR cycles using external control (single band amplification reaction, SBA), the different concentrations of gene specific and control primer pairs required to have a relative amplification of the control and target products in a DA reaction was then calculated. The concentration of the primer pairs was varied until there was a relatively similar amplification of the esterase D and C16*orf*5 PCR products. The amount of primers used for esterase D

and C16*orf*5 in duplex amplification were 15 ng and 250 ng respectively. Finally, the effects of increasing number of cycles on the PCR amplification of the C16*orf*5 and the esterase D PCR products in the patient with t(1;16) and a normal control was also studied. The RT-PCR products were analyzed on 1.5% agarose gels with 0.5 ug of ethidium bromide staining in 1 X TBE. Quantitative analysis was performed using the NIH Image 1.54 software (the/pub/nih-image directory) to calculate the arbitrary units of the signal intensity of the respective RT-PCR amplified products.

The following gene specific primers were used in the RT-PCR experiments: Exon 1- Forward primer, 5'-GAAGCAGCGAAGATGTCCAGC-3', Reverse primer, 5'-GGAGTGGCATGCCTGGAG-3'; Exon 2- Forward primer 5'-CTCCAGGCATGCCACTGC-3', Reverse primer, 5' – GGAGGGTAGAAACCCGGAGG-3'. The primers used for RT-PCR analysis were designed as described in section 2.3.5.1. The sequences for the esterase D (positive control) forward and reverse primers used are presented in section 2.3.8.

7.2.4 Mutation detection

For mutation screening of the coding regions of the C1*orf*5 gene, genomic DNA from 100 female patients diagnosed with non-specific mental retardation was submitted to PCR amplification. These samples were kindly provided by Dr. Agi K. Gedeon, Department of Cytogenetics, WCH, Adelaide. Primers were designed flanking each exon at ~50-80 bp distal and proximal in the introns as described in section 2.3.5.1. The primers used for mutation detection analysis were synthesized with the HEX fluorescent label. Subsequently, SSCP analysis was performed by Joanna Crawford (Department of Cytogenetics, WCH, Adelaide). Briefly, PCR

amplification was carried out with initial denaturation for 5 minutes at 95°C, followed by amplifications for 10 cycles under the following conditions: 94°C for 1 minute, 60°C for 1.5 minute, then 72°C for 1.5 minute. This was then followed by 25 cycles of 90°C, for 1 minute, 55°C for 1.5 minute and 72°C for 1.5 minute, with a final extension at 72°C for a further 10 minute. The completed PCR reaction was then diluted 1 in 3 with formamide loading dye, denatured at 94°C for 2 minute, placed on ice prior to loading 1 µl of the final volume onto a 4.5% non denaturing polyacrylamide gel. SSCP analysis was performed on the GEL-SCAN 200 real time gel system (Corbett Research, Australia). The bands showing altered mobility were then reamplified and sequenced using the original PCR primers (minus HEX) by the candidate.

7.3 Results

7.3.1 C16orf5 expression in affected male with non-specific MR

The chromosome 1 breakpoint of the t(1;16) was shown to be within the pericentromeric heterochromatin (section, 5.3.2). It was possible that the C16*orf*5 expression pattern was decreased due to a position effect of the adjacent translocated chromosome 1 heterochromatin. In view of this hypothesis, semi-quantitative RT-PCR analysis involving C16*orf*5 was performed to study its expression in the patient with t(1;16) as compared to a normal individual.

7.3.1.1 RT-PCR analysis of C16orf5 using RNA from lymphoblastoid cells

In order to study the expression pattern of the C16*orf*5 gene in the patient as compared to a normal individual, RNA was extracted from the established lymphoblastoid cell lines of the patient and an individual with a normal karyotype (section 7.2.1).

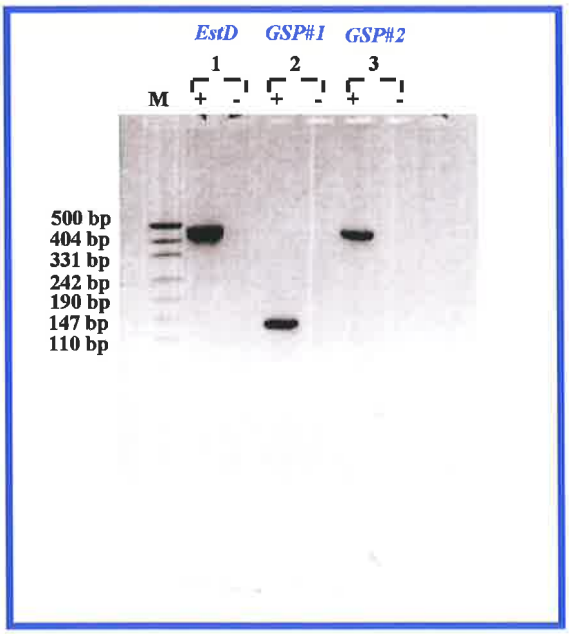


Figure 7.1 RT PCR analysis of C16*orf*5 gene using lymphoblastoid RNA. Samples: + indicates lanes with reverse transcriptase enzyme and - indicates lanes without reverse transcriptase enzyme. Esterase D (EstD) primers were used as a positive control in the experiment. Both the gene specific (GSP) primer pairs amplified RT-PCR products from the lymphoblastoid RNA. The product sizes of pUC19 (M) marker bands are indicated.

Initially, RT-PCR analysis was performed using the C16*orf*5 gene specific primer pairs (section 7.2.2) and RNA from lymphoblastoid cells to ensure that the gene is expressed in these cells. The RT-PCR results showed that C16*orf*5 gene was expressed in lymphoblastoid cells (Figure 7.1). Esterase D primers (section 2.3.8) were used as a positive control in the experiment.

7.3.1.2 Semi-quantitative RT-PCR analysis of C16*orf*5 gene

The C16orf5 gene was the closest identified gene to the 16p13.3 breakpoint of the t(1;16) and was highly expressed in brain. Therefore, it was targeted for quantitative RT-PCR analysis to study its expression in the patient as compared to the normal individual. An example of the RT-PCR analysis with exon 1 and esterase D primers in duplex amplification is shown in Figure 7.2 A. The results shown in Figure 7.2 A were further confirmed by performing the same experiment at different number of PCR cycles (Figure 7.2 B). An example of the gene specific and esterase D RT-PCR products from a normal control and the patient scanned using NIH Image 1.54 software is depicted in Figure 7.2 C. The arbitrary relative values obtained for the signal intensity of the esterase D and C16*orf*5 RT-PCR products from the agarose gel shown in Figure 7.2 B are depicted in Table 7.1. The arbitrary relative values obtained for the signal intensity of the respective products suggested that the expression of C16*orf*5 was reduced by average of 14% in the lymphoblastoid cell line from the patient compared to the normal control cell line. This was less than 50% reduction expected from the position effect. However, it was not possible to study the expression of C16*orf*5 in the appropriate tissue (brain) due to the unavailability of RNA samples. It is possible that the observed level of reduced expression of this brain specific transcript in the patient is a consequence of normal human variation.

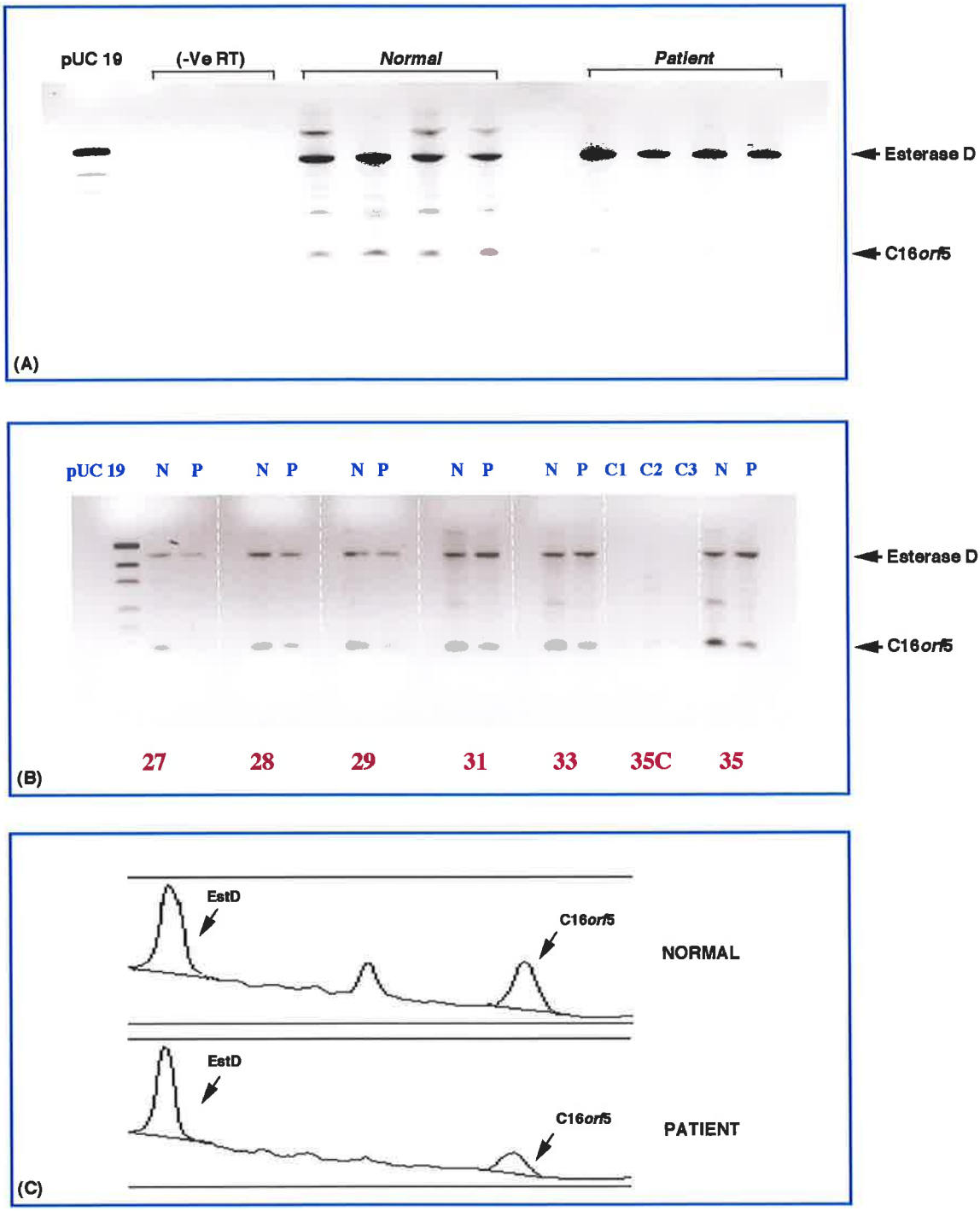


Figure 7.2 C16*orf*5 gene expression analysis in a normal individual and the patient with mental retardation. **(A)** The agarose gel showing Esterase D and C16*orf*5 RT-PCR products. PCR for each sample of normal and patient was performed in quadruplicate. Esterase D primers were used as a control. **(B)** The gel showing the effect of increasing number of PCR cycles on the amplification of Esterase D and C16*orf*5 RT-PCR products. The PCR was performed in the set of six for each of the normal (N) and patient (P) sample. Tubes for each set were taken out after 27, 28, 29, 31, 33 and 35 PCR cycles. The lanes at 35 C are controls with no DNA. **(C)** Quantitation of the results shown in 5B. The gel was scanned using the NIH Image 1.54 software. An example of scanned RT-PCR products (upper panel-> normal individual; lower panel-> affected male). Arrows indicate the individual peakes for the Esterase D and C16*orf*5 products.

Table 7.1 Quantitation of the results shown in Figure 7.2 B. This Table depicts the arbitrary relative values obtained for the signal intensity of the esterase D and C16*orf*5 RT-PCR products amplified from the template DNA of the normal control and patient with t(1;16) using NIH Image 1.54 software.

PCR cycles	Esterase D expression (arbitrary units)		C16*orf*5 expression (arbitrary units)		% reduction of C16*orf*5 expression in patient
	Normal	*Patient*	*Normal*	*Patient*	
27	668	642	379	327	6%
28	710	700	517	410	14%
29	710	710	534	408	18%
31	744	730	632	500	16%
33	748	732	647	501	18%
35	782	744	700	595	10%

Arbitrary relative value for esterase D expression in the normal lymphoblastoid cell line was taken to be 100%.

7.3.2 Mutation analysis involving C16orf5 gene

Mutation analysis was carried out to further investigate the possible role of C16*orf*5 gene in non-specific mental retardation in the human population.

7.3.2.1 Intron exon boundary organization of C16*orf*5 gene

The partial genomic structure of the gene was determined (section, 3.3.5), which identified at least four tentative exonic boundaries for the C16*orf*5 gene. However, the sequences of the introns were not known at this stage. Later, DNA sequence database searches performed on May 11, 2000 with C16*orf*5 cDNA sequence (Figure 3.4) identified the working draft sequence of the BAC 35P16 (RPCI-11, BAC library) that was available in GenBank under the accession number, AC007606 as two ordered pieces.

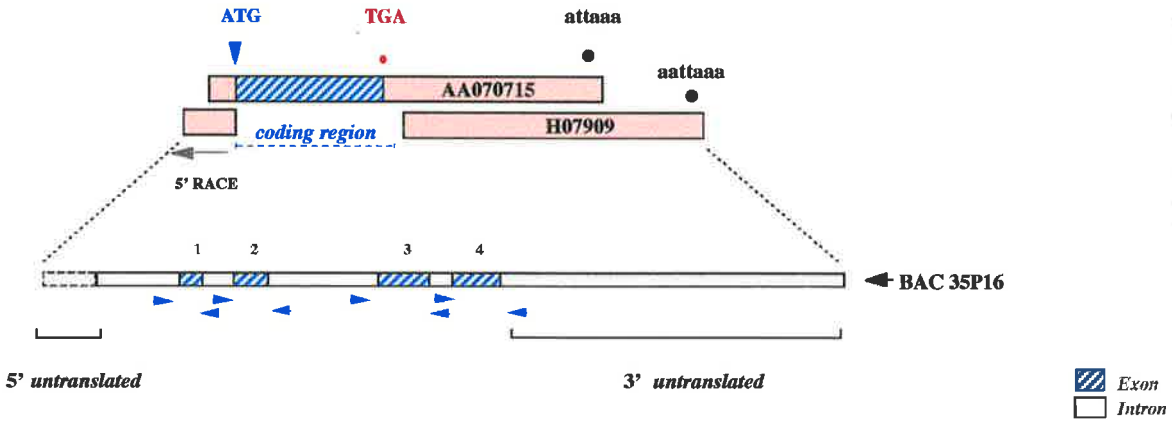


Figure 7.3 C16*orf*5 gene structure. The genomic structure was established by comparison of C16*orf*5 cDNA sequence with BAC 35P16. The gene spans over 4.1 kb of the genomic DNA. Arrows indicate the location of primers designed for SSCP analysis and DNA sequencing.

Identification of the BAC 35P16 spanning the C16*orf*5 gene enabled the determination of the intron–exon organization and sequence of the boundaries of the C16*orf*5 gene (Figure 7.3, Table 7.2). The intron/exon splice sites junctions conform to the ag/gt rule (Shapiro and Senapathy, 1987). This allowed the design of primers flanking each exon at ~50-80 bp distal and proximal to the coding sequence to search for mutations in the open reading frame of the C16*orf*5 gene.

Table 7.2 Exon-Intron boundaries of C16*orf*5 gene.

Exon #	Acceptor site	Donar site	Exon size (bp)	Intron size (bp)
	5'UTR	CGGATTTGAG/**gt**aagaaatc	281	430
1	ccatgtcc**ag**/GAGAAGCAGC	CCCACCCCAG/**gt**aggggtc	99	194
2	ctctgtcc**ag**/GCCGTTCCTC	ATGCCTCCGG/**gt**gagtgggg	156	631
3	gccattcc**ag**/GTTTCTACCC	GCTTCATGGG/**gt**aggtggtg	274	107
4	tctgattc**ag**/ATGTGATCTG	GAGCCGC**TGA***/3'UTR	269**TGA***	3'UTR (1654) **aataaa**

The splice sites junction sequence conforms to the ag/gt rule (Shapiro and Senapathy, 1987).

7.3.2.2 Mutation analysis

DNA from 100 female patients diagnosed with non-specific mental retardation was used for the mutation screening of the coding regions of the C1*orf*5 gene. The primers used for mutation detection are shown in Table 7.3. A single nucleotide polymorphism was detected in three patients in exon 4 of the C16*orf*5 gene at nucleotide position 983 of the C16*orf*5 cDNA sequence (Accession number, AF131218, Figure 3.4). This resulted in a substitution of the nucleotide (C->T), in

the third position of a codon (TGC) for the cysteine residue. This does not result in an amino acid change. Another nucleotide substitution was observed in the third patient in the flanking intronic sequence of exon 4 (C->T), 23 bp distal to the coding sequence. No mutations were detected in the coding sequence of the C16*orf*5 gene that changed the aminoacid sequence in any of the samples examined. It was concluded that mutation in the C16*orf*5 gene is unlikely to be a cause of sporadic cases of mental retardation.

Table 7.3 Primers used to amplify the coding regions of C16*orf*5 for mutation detection.

Primer name	Forward primer (5'-3')	Reverse primer (5'-3')	PCR product size (bp)
Exon#1/ F/R	TGGGCATCACCACTTAGAGA	CCACAGGGCTGAAGGAGA	247
Exon#2/ F/R	CAGACAGCAGCCAGGAGTTC	TCTGTGTCCCAAAACAGTGG	294
Exon#3/ F/R	CAGAGGCAGCCAGTTGACC	CGCCAATCTCTTTGTCTGGT	403
Exon#4/ F/R	CACTGACCCTTGGCCTTAAA	GGTCAACTGGCTGCCTCTG	384

Each primer set was prefixed with "kav".

7.3.3 Identification of A2BP1 gene (ataxin–2 binding protein 1).

A recent search (dated: November 11, 2000) of the sequence databases using the BLAST algorithm with the sequence of BAC 375G12, which is distal to the 16p13.3 breakpoint of the t(1;16) (Figure 5.10), detected the 5' UTR of the novel protein A2BP1 (ataxin–2 binding protein 1). The A2BP1 cDNA sequence (2279 bp) became available in the GenBank on November 2, 2000 under the accession number NM_018723.

7.3.3.1 Physical mapping of A2BP1 with respect to 16p13.3 breakpoints of the t(1;16) and t(14;16)

The cDNA sequence of A2BP1 (2279 bp, accession number NM_018723) was used to detect homologous sequences in the non-redundant and htgs databases of NCBI. These sequence analyses identified seven BAC clones (375G12, 167B4, 185J20, 545E8, 192K18, 2357P14, 26O3) with matches to exons of the A2BP1 sequence. These BACs had all been previously mapped in the vicinity of the t(1;16) and t(14;16) chromosome 16 breakpoints (Figures 5.10 and 6.10). The BAC 167B4 (RPCI-11), which is not shown in the Figures 5.10 and 6.10 was subsequently mapped in the region by PCR analysis of DNA markers D16S3232 (WI-4274), D16S3128 and T53018 (Figure 4.7). The marker D16S3232 (WI-4274) is contained within the BAC 167B4. Pulsed-field gel electrophoresis of restricted DNA using the enzymes *Not*I and *Eag*I estimated the size of the BAC 167B4 to be 174 kb.

The comparisons of A2BP1 cDNA sequence with the genomic sequences of the above BACs allowed the determination of 16 exons and the relevant exon/intron boundaries, Table 7.4. A2BP1 spans at least 1.6 Mb of the genomic DNA and the chromosome 16p13.3 breakpoints of the t(1;16) and t(14;16) are located in introns 1 and 3 respectively of the A2BP1 gene. The positions of the 16 A2BP1 exons with respect to the BACs 375G12, 167B4, 185J20, 545E8, 192K18, 2357P14, 26O3 and the 16p13.3 breakpoints of the t(1;16) and t(14;16) are shown in Figure 7.4.

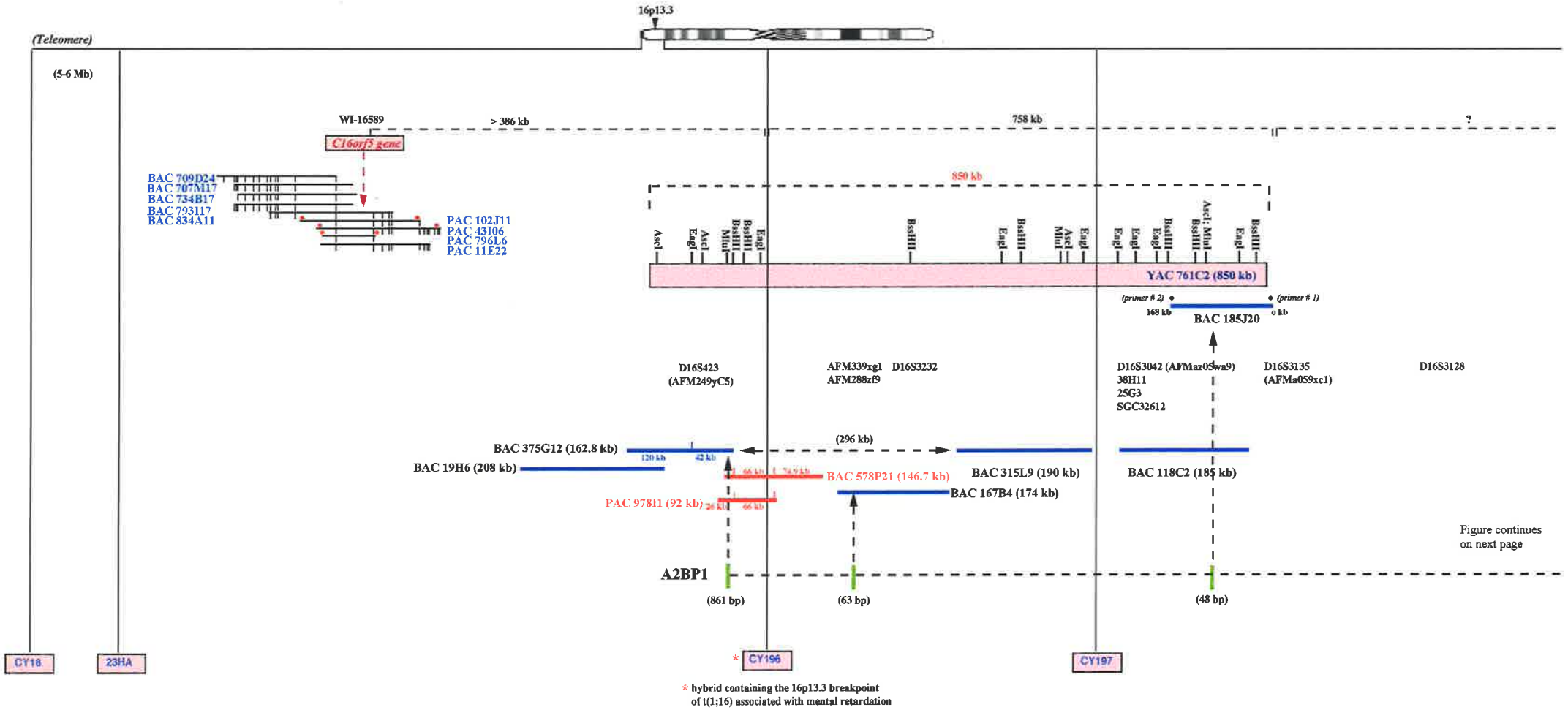
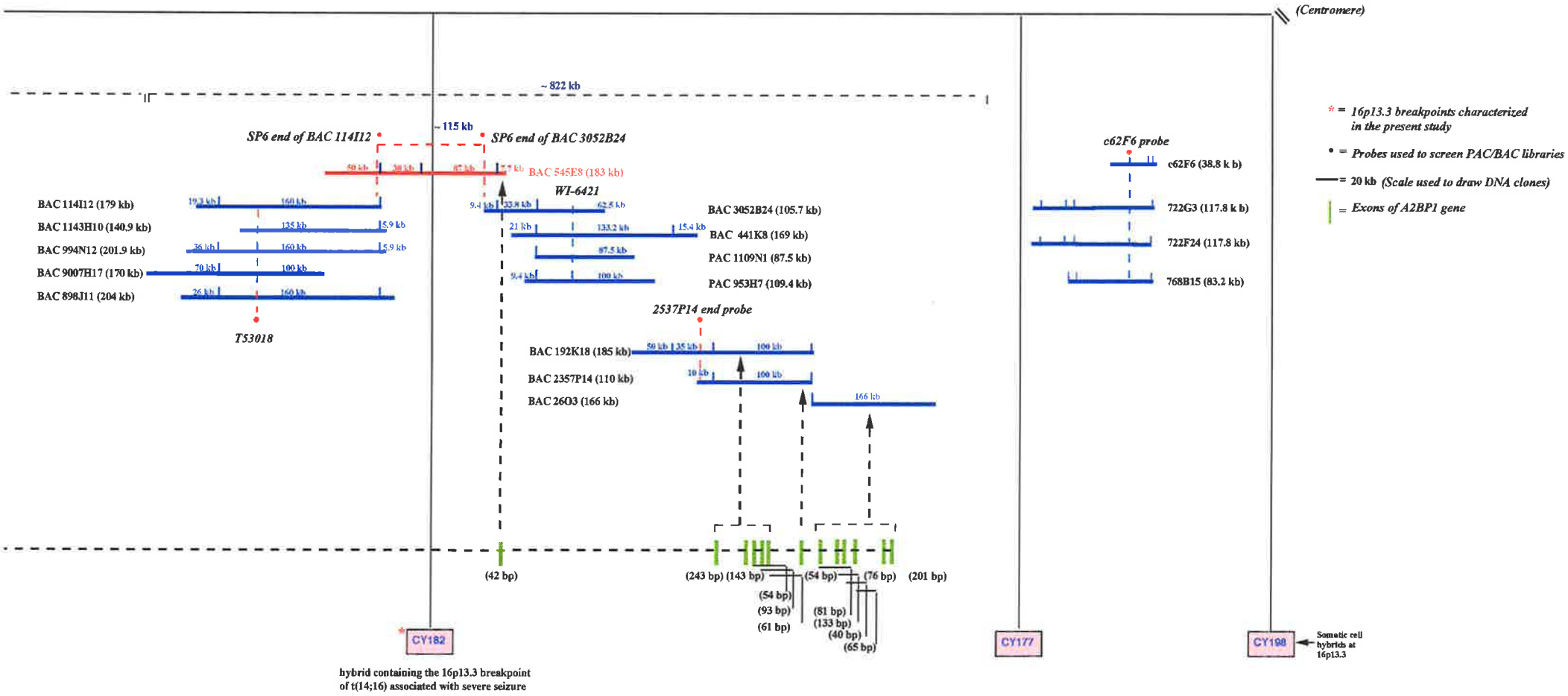


Figure 7.4 A2BP1 gene is disrupted by the 16p13.3 translocation breakpoints of the t(1;16) and t(14;16) characterized in the present study. The positions of A2BP1 exons with respect to the 16p13.3 breakpoints are shown. 16p13.3 breakpoints of both the translocations t(1;16) and t(14;16) reside in the introns of A2BP1 gene. The results from Figures 4.7, 5.10 and 6.10 are included in order to show the position of A2BP1 gene with respect to the DNA clones identified in the vicnity of the 16p13.3 breakpoints. The Figure is presented in two pages. The position of A2BP1 gene with respect to BAC clones identified to contain A2BP1 exons by BLAST search analysis of A2BP1 cDNA sequence is shown by the black dashed lines with arrows at the start. Sizes of the exons are indicated. The A2BP1 exon one was found to be located on BAC 375G12 by BLAST search analysis at a distance of 4 kb from the 66 kb fragment of PAC 978I1 that contains the 16p13.3 breakpoint of the t(1;16). The length of the YAC 761C2 proximal to the 16p13.3 breakpoint of the t(1;16) is 758 kb. The length of the contig spanning the 16p13.3 breakpoint of the t(14;16) is 822 kb. (Figure legend continues on next page).



These distances when added reveal that the A2BP1 gene spans at least 1.6 Mb of genomic DNA. The DNA clones that are shown in red were identifed to span the 16p13.3 breakpoints of the t(1;16) and t(14;16). The distance between the BAC 185J20 and the contig spanning the marker T53018, which was localised in the hybrid interval defined by CY197 and CY182 is not known. The BAC 167B4 is not shown in the Figure 4.7 because it was identified later in the course of study during an attempt to clone the 16p13.3 translocation breakpoint of the t(14;16). The BAC was found to be positive for the marker D16S3232 by PCR analysis. The order of the breakpoints of the hybrids at 16p13.3 was determined to be pter-23HA-CY196-CY197-CY182-CY177-CY198.

Table 7.4 Exon-Intron boundaries of A2BP1 gene.

Exon #	Acceptor site	Donar site	Exon size
1	5'UTR	GGGAGTTCTAG/**gt**aagtccag	860
2	tcttctttt**ag**/GAAACTGGTC	TCAGCATTCA/**gt**aagtgcaa	62
3	ctcttctc**ag**/GATATCAAAG	AGACAGAAAG/**gt**gagtcaat	48
4	ttctttct**ag**/GTTTCAAGAC	GCAGCTAAGG/**gt**aggtgcac	42
5	gatttttc**ag**/GGTAATCAGG	CACCGCCACA/**gt**aagtggac	243
6	gttcttttag/CAGACAGATG	AAATGTTTGGT/**gt**aagtatca	143
7	ttctttgc**ag**/CAATTTGGTA	AGGCTCAAAG/**gt**aagcaact	54
8	ctgcatgc**ag**/GGATTTGGTT	TAAAATCGAG/**gt**gcatgttc	93
9	atgtacat**ag**/GTAAATAATG	TATACAAATG/**gt**aagtagag	61
10	ttgttttaag/GCTGGAAATT	TTCTATGCAG/**gt**acagagtt	54
11	tctttcgt**ag**/GCACGGTCCT	ACTTCTGCAA/**gt**aagcccac	81
12	tgccccgc**ag**/TGCCAGGCTT	CCTACGGCGG/**gt**aagtgggg	133
13	acccttgc**ag**/TGTTGTTTAC	AGACATTTAT/**gt**aagtattc	40
14	tgtgttttag/GGTGGTTATG	ACAGTGACAG/**gt**aagggtca	65
15	ttcctttc**ag**/TTACGGACGA	TGGTGCCATG/**gt**gagtacaa	76
16	tttgtttc**ag**/AATGCTTTTG	3'UTR	201

The splice sites junction sequence conforms to the ag/gt rule (Shapiro and Senapathy, 1987).

7.4 Discussion.

Because the chromosome 1 breakpoint of the t(1;16) was shown to be within the pericentromeric heterochromatin (Figure 5.1 A), it was hypothesized that the phenotype observed in the patient might be a result of a position effect on the expression of the C16*orf*5 gene due to the adjacent translocated chromosome 1 heterochromatin. The C16orf5 gene was a potential candidate because at this time it was the closest identified gene to the 16p13.3 breakpoint of the t(1;16) and was highly expressed in brain. Therefore, semi-quantitative RT-PCR analysis of

C16orf5 gene was performed to study its expression in the patient as compared to the normal individual using the lymphoblastoid RNA. Semi-quantitative RT-PCR analysis showed that the expression of C16*orf*5 was reduced in a lymphoblastoid cell line from the patient compared to a normal cell line. However, it was not possible to further study the expression of C16*orf*5 in the appropriate tissue (brain) due to the unavailability of RNA samples. It is possible that the reduced expression of this brain specific transcript in the patient is a consequence of normal human variation.

Therefore to further evaluate the role of C16*orf*5 gene in mental retardation mutation analysis of this gene was carried out. Identification of BAC 35P16 containing the C16*orf*5 cDNA sequence allowed the determination of sequence of the intron/exon boundaries. This enabled the design of primers flanking each exon at ~50-80 bp distal and proximal to the coding sequence. These primers were used to search for mutations in the open reading frame of C16*orf*5 gene by SSCP in 100 female patients with non-specific mental retardation. However, no coding mutations that altered the protein sequence were seen. This suggests that C16*orf*5 gene is unlikely to be involved in mental retardation.

Subsequent to this work being completed, the gene A2BP1 was identified by sequence homology to the sequence of BAC 375G12, which was located distal to the 16p13.3 breakpoint of the t(1;16). A2BP1 is highly expressed in brain and muscle and binds to the C-terminus of ataxin-2 (Shibata et al., 2000). An expansion of a polyglutamine tract in ataxin-2 is known to be responsible for spinocerebellar ataxia type 2 (Huynh et al., 1999). It is suggested that binding of A2BP1 to ataxin-2 may be involved in the cell type-specific neuronal death seen

in SCA2. The exact role of the ataxin-2/A2BP1 protein complex is not resolved. However, it is suggested that if RNA binding domains in both proteins are functional, ataxin-2/A2BP1 complex may have a role in RNA transport in neurons (Shibata et al., 2000).

Nucleotide database searches using the cDNA sequence of A2BP1 identified homologies to the BAC clones 375G12, 167B4, 185J20, 545E8, 192K18, 2357P14 and 26O3. These were used for the cloning of the 16p13.3 translocation breakpoints of the t(1;16) and t(14;16) confirming the location of this gene. The comparison of the A2BP1 cDNA sequence with the genomic sequences for these BACs showed that the chromosome 16p13.3 breakpoints of the t(1;16) and t(14;16) reside within introns of the A2BP1 gene. The established physical map of the breakpoints region, together with the cDNA sequence of A2BP1, suggest that A2BP1 gene consists of 16 exons and spans at least 1.6 Mb (Figure 7.4) of the genomic DNA.

Because A2BP1 is predominantly expressed in brain and muscle tissues, ataxin-2/A2BP1 complex might have a role to play in RNA transport in neurons, and the A2BP1 transcript is directly disrupted by the 16p13.3 breakpoints of both *de novo* chromosomal rearrangements, it is highly probable that A2BP1 is involved in the clinical phenotypes of the two patients. Identification of this disrupted gene provided a candidate for further mutation and expression analyses (currently underway in the Department of Cytogenetics and Molecular Genetics, WCH, Adelaide) in other affected patients to assess the involvement of A2BP1 in sporadic autosomal mental retardation and/or epilepsy.

Conclusion

Chapter 8

Mental retardation and/or epilepsy account for a majority of childhood neurological disorders. A major priority of neuroscience research today is the identification of genes related to brain disorders. Identification of families with mentally retarded males has led to the mapping and cloning of a number of X-linked genes for mental retardation. However, the etiologies of most non-specific autosomal MR cases are unknown. The identification of equivalent genes on autosomes is more problematical because autosomal recessive genes related to such disorders are unlikely to be identified except in specialised circumstances (example, defined populations with founder effects, families with inbreeding). Autosomal dominant forms may also be difficult to genetically map due to the likely presence of genetic heterogeneity if the phenotype is relatively non-specific. In addition, individual pedigrees are unlikely to be large when affected individuals have a relatively severe phenotype.

There are several published reports in which individuals with balanced *de novo* translocations involving auotsomes have had MR of varying degrees, however, no autosomal gene for a non-specific form of mental retardation has yet been identified. Another approach being practiced is to identify families associated with nonspecific mental retardation and map the gene by linkage analysis. There is only a single published report suggesting the localisation of a gene for autosomal NSMR to 3p25-pter (Higgins et al., 2000). At this time, the gene responsible for NSMR in this family has not yet been identified. Mutations in some ion channel genes (refer, section 1.8.2.2) are now shown to be causes of epilepsy, but due to genetic heterogeneity the molecular mechanisms involved in the clinical phenotypes are not completely understood. Hence, to better understand the genetics underlying mental retardation and epileptic seizures, it is

necessary to identify more disease genes associated with the functions of brain by exploiting any opportunity to clone such genes.

Linkage analysis has been successful in mapping various gene loci for MR and epilepsy. But this technique requires a single large family to accurately map the disease-related gene, which is not always possible. A possible alternative is the positional cloning of chromosomal rearrangements associated with clinical phenotypes (Villard et al., 1999). However, the occurrence of such cases is very rare, and therefore when identified, it is important to explore them for the identification of disease genes. Two such patients were reported at the Department of Cytogenetics, WCH, Adelaide, with balanced *de novo* translocations t(1:16) and t(14;16) associated with mental retardation and epilepsy respectively. The mouse/human somatic cell hybrids isolated from the cell lines of these two patients containing the der(16)t(1;16) and the der(16)t(14;16) were designated as CY196 (Callen et al., 1990b) and CY182 (Callen et al., 1995) respectively. These two chromosome 16 breakpoints were included in the chromosome 16 somatic cell hybrid panel that was used to physically map chromosome 16 in detail with an average resolution of 1-Mb (Callen et al., 1992, 1995). The translocation breakpoints in the two patients were localised in close proximity on chromosome 16 at 16p13.3 (Callen, 1995). It was hypothesised that mental retardation and epilepsy in two patients with *de novo* balanced translocations is due to a common gene defect at or near the 16p13.3-translocation breakpoints. These two translocation cases provided a unique opportunity to clone the gene related to brain function in the two patients. Studies involving patients with translocations associated with a particular disease has earlier led to the cloning of genes like NF-

1 (neurofibromatosis-1) (Ledbetter et al., 1989), and Duchenne muscular dystrophy (Koenig et al., 1987).

In the current study, the positional cloning strategy was chosen as an approach to identify the disease gene responsible for the phenotype observed in the two patients. The major aims of the study were to clone the 16p13.3 translocation breakpoints of the t(1;16) and t(14;16) using BAC/PAC clones, subsequently sequence the genomic insert of interest and then analyse the sequence data for the presence ESTs and genes using publicly available sequence comparison tools such as BLAST (Altschul et al., 1990, 1994) and FASTA (Wilbur and Lipman, 1983) at NCBI to identify if there was any gene disrupted at the breakpoints. This was chosen as the method of choice because by the time the present project was started in 1997, the identification of human genes was revolutionised by the advent of large-scale cDNA sequencing (Brenner, 1990) and generation of ESTs (Adams et al., 1991) as a part of the HGP. By 1995, the number of the ESTs in the GeneBank was increased by thousands and a human gene catalogue was created. Adams et al., (1995) assembled 118,406 ESTs, present at that time in dbEST, into 29,599 distinct assemblies (contigs), which were notated as THCs. At NCBI, ESTs representing single gene were also assembled in cluster in the UniGene database (Schuler et al., 1996). The version of the human gene map published in 1998 by Bentley and colleagues reported the identification and localisation of 18,703 UniGene clusters. The EST strategy allowed rapid identification of expressed genes by sequence analysis, which transformed the previous tedious laboratory based methodology such as direct cDNA selection and exon trapping into a largely in silico based approach. The most recent developing approach for gene identification is cDNA microarray technology or DNA chip technology,

which allows expression monitoring of hundreds and thousands of genes simultaneously, and provides a format to identify genes as well as changes in their activity (Heller et al., 1997). The ultimate goal of this technology is to develop arrays that contain every gene in the human genome against which mRNA expression can be assessed. However, this technology has only been recently developed and was not available in the Department during the course of the present study and therefore was not used in this thesis. The efforts of HGP and Celera Genomics in sequencing the human genome will ultimately allow the identification of all human genes. However, until an array with a complete set of human genes is available, positional cloning approaches will still be required for the identification of particular disease genes.

At the beginning of the project in 1997, there was not a DNA marker or clone mapping within the hybrid interval defined by the CY196 and CY182 breakpoints. Therefore, it was necessary to prepare a detailed physical map of this region at 16p13.3. This preliminary mapping identified a novel proline rich gene C16*orf*5, highly expressed in the brain (Chapter3). To localize this gene in more detail it was necessary to generate a detailed physical map of the region. A long-range YAC restriction map of the breakpoints region was prepared using YAC 761C2 and the BACs 375G12, 315L9 and 118C2 (Chapter 4). YAC maps have made important contributions to the identification of many candidate disease genes including Huntington disease (Rommens et al., 1993) and the OATL1 (ornithine - δ-aminotransferase) gene associated with retinal disorders (Geraghty et al., 1993). However, the long-range YAC restriction map of the breakpoints region revealed that C16*orf*5 gene was in fact distal to the 16p13.3 breakpoint of the t(1;16) and was not disrupted by either of breakpoints at 16p13.3. This map also separated the

two translocation breakpoints of the t(1;16) and t(14;16) by at least a distance of 758 kb, making it less likely that the same gene was involved in the observed phenotypes of the two patients. Therefore, to further refine the two-translocation breakpoints and in order to identify if any transcript is directly disrupted by these 16p13.3 breakpoints, a BAC/PAC contig was constructed (Chapters 5 and 6).

The search for the gene responsible for the observed phenotype in the patient with t(1;16) was focussed on chromosome 16 because the chromosome 1 breakpoint was shown to be within 1q heterochromatin. The YAC restriction map of the breakpoints region at 16p13.3 (Chapter 4) narrowed the 16p13.3 breakpoint of the t(1;16) to the 296 kb gap between the BACs 375G12 and 315L9. BAC 375G12 sequence identified an overlapping BAC 578P21 that was positioned in the 296 kb gap. This clone was found to span the 16p13.3 breakpoint by FISH. The sequence of this clone was not available in the GenBank at that time and the insert size as estimated by PFGE was 147 kb. Due to budgetary restraints it was necessary to identify a smaller DNA clone, which could be more economically sequenced. Additional clones were identified by screening high-density filters of genomic PAC clones. From this effort the PAC 978I1 was found to span the 16p13.3 breakpoint of the t(1;16) from FISH and PCR analysis. DNA of this spanning PAC was sequenced with 1.5X coverage to allow the possible identification of a transcript, which was disrupted at the breakpoint. Analysis of this PAC sequence using the dbEST database at NCBI indicated that the region surrounding the breakpoint is repetitive with very few matches to expressed sequences (ESTs). These ESTs and exons that were identified were generally present as singletons, did not contain poly A signals, and were contiguous with genomic sequence. Based on this information it was unlikely that these ESTs were derived from a

genuine transcript. Furthermore, expression studies with some of these ESTs matches also failed to identify evidence of an expressed transcript. The PAC 978I1 sequence was also analysed for the presence of exons using NIX program. NIX is a WWW tool available at UK HGMP Resource Centre to allow the results of many analysis programs to be identified simultaneously. The analysis programs run includes GRAIL, Fex, Hexon, MZEF, Genemark, Genefinder, Fgene, BLAST, Ployah, RepeatMasker and tRNA scan. There were two regions on the PAC 978I1 sequence, where exons were predicted by four different programs. However, expression analyses with these computer-predicted exons failed to identify any evidence of an expressed transcript. Although, there are many useful computer tools available to identify transcribed or other interesting regions in the human genome none of these are 100% accurate (Claverie, 1997). It was concluded from these analyses that the chromosomal rearrangement either resides in a large intron of a causative gene, the observed phenotype was the result of a position effect of the chromosome 1 heterochromatin on a nearby gene or the association of translocation and clinical abnormalities was fortuitous.

The 16p13.3 breakpoint of the t(14;16) was also successfully cloned. This was facilitated by the increasing availability of human genome sequence in the NCBI database as a consequence of a rapid progress of the HGP. A BAC 545E8 spanned the 16p13.3 breakpoint of the t(14;16) and sequence was available for this clone in GenBank. In silico analysis identified EST matches that were present as singletons. On the basis of experience of expression results with such singleton EST matches for the PAC 978I1 sequence it was decided not to proceed with further studies of these ESTs. Further, FISH results of BAC 545E8 in the patient with the t(14;16) suggested that the translocation in this patient was also

associated with a paracentric inversion with a coincident breakpoint at 16p13.3, t(14;16)(p32;p13.3)inv16(p13.3p12.1). This was confirmed by FISH analysis using a series of BAC/PAC DNA clones mapping proximal and distal to the 16p13.3 breakpoint of the t(14;16). The unexpected three break complex chromosomal rearrangement complicated the further positional cloning approach.

One hypothesis was that the observed phenotype in the patient with t(1;16) might be the result of a position effect of the chromosome 1 heterochromatin. Juxtaposition of a gene to a region of heterochromatin through chromosomal rearrangements can result in the alteration of gene expression (Karpen, 1994). Accordingly, expression and mutation analyses (Chapter 7) of the C16orf5 gene, the closest identified gene to the 16p13.3 breakpoint of this translocation were carried out. The data generated suggested that it was unlikely that C16orf5 is a gene frequently involved in sporadic mental retardation. However, the C16*orf*5 gene itself is of interest because of the unusually high content of proline residues (40% over 104 residues) at the N-terminus of the protein and 14 cysteine residues at the C-terminus of the protein. The predicted nuclear localization and abundant brain specific expression suggest that C16*orf*5 is likely to have an important role in the development and function of neuronal cells.

During the final stages of this project a gene A2BP1 was identified and sequenced (Shibata et al., 2000). Sequence analysis with BAC 375G12, which was located distal to the 16p13.3 breakpoint of the t(1;16), showed homology to the 5'UTR of this gene. A2BP1 is highly expressed in brain and muscle and binds to the C-terminus of ataxin-2 (Shibata et al., 2000), which is the gene product of the human SCA2 gene that is known to cause Spinocerebellar ataxia type 2 (Huynh et al.,

1999). Additional analysis involving the A2BP1 cDNA sequence together with the physical map of the breakpoints region generated in the present study revealed that the gene spans at least 1.6 Mb of genomic DNA and both the chromosome 16p13.3 breakpoints of the t(1;16) and t(14;16) were found to be located in introns of this gene. The sequenced regions of the PAC 978I1 and the BACs 441K8 and 545E8 that were used for expression analyses reside in introns of the A2BP1 gene. This result explains the failure of transcript identification by RT-PCR and Northern analyses in the region. The A2BP1 transcript is directly disrupted by the 16p13.3 breakpoints of both *de novo* chromosomal rearrangements, the gene is predominantly expressed in brain and muscle tissues, and the ataxin-2/A2BP1 complex might have a role to play in RNA transport in neurons (Shibata et al., 2000). Therefore, it is highly probable that A2BP1 is involved in the clinical phenotypes of the two patients. However, it may also be speculated that if any candidate genes are present at or near the other two autosomal breakpoints (inversion 16p12.1 and 14p32) of the t(14;16), they could contribute to the severe epilepsy phenotype observed in this patient.

The physical map data presented in the present study refined the order of the chromosome breakpoints of the hybrids at 16p13.3 to pter-23HA-CY196-CY197-CY182-CY177-CY198-CY168 from that presented in Callen et al., (1995). This placed the 16p13.3 breakpoint contained in CY197 (Callen et al., 1990b) between the 16p13.3 breakpoints of the t(1;16) and t(14;16) contained in C196 and CY182 respectively. The CY197 hybrid was generated in 1990 using the A9 mouse cell line and a human fibroblast cell line containing the chromosome 16 translocation 46,XY,t(13;16)(q13.3;p13.3). This cell line was kindly provided by Dr R.J Desnick, Mount Sinai Medical Centre, New York, USA and originated from a

chronic villus biopsy at prenatal diagnosis. The translocation was familial and the phenotype of the patients was reported to be normal. Because the present study has identified that the A2BP1 gene is disrupted by this translocation, the clinicians associated with the case have been contacted to review the patients' histories. Unfortunately, they are lost to follow up. It is also possible that the derivative chromosome 16 in the hybrid cell line CY197 has undergone culture-induced deletion. This supposes that the original 16p13.3 breakpoint of the t(13;16) is distal to the chromosome breakpoint in CY196 but subsequently has undergone rearrangement in the somatic cell hybrid resulting in loss of a region of 16p13.3. Work is in progress to determine the integrity of the chromosome 16 in this cell line in the Department of Cytogenetics and Molecular Genetics, WCH, Adelaide.

The current study enabled to determine that the A2BP1 gene was disrupted by the two translocations, t(1;16) and t(14;16). These translocations were *de novo* and the carrier patients had either severe intellectual retardation or severe seizures. The original hypothesis, that mental retardation and epilepsy in these two patients is the result of disruption of the same gene at 16p13.3 was supported. The disrupted gene A2BP1 is an excellent candidate since it is predominantly expressed in brain and muscle tissues. The work presented in this thesis provides the basis for further work to determine if A2BP1 is involved in sporadic autosomal mental retardation and/or epilepsy. Such studies are currently underway in the Department of Cytogenetics and Molecular Genetics, WCH, Adelaide. Identification of the gene related to mental retardation and/or epilepsy will further help in better understanding of the molecular basis of neuronal mechanisms underlying these disorders.

REFERENCES

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos, MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR and Venter JC (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. Science **252**:1651-1656.

Adams MD, Kerlavage AR and Fleischmann RD Fuldner RA, Bult CJ, Lee NH, Kirkness EF, Weinstock KG, Gocayne JD, White O, Sutton G, Blake JA, Brandon RC, Chiu MW, Clayton RA, Cline RT, Cotton MD, Earle-Huges J, Fine LD, FitzGerald LM, FitzHugh WM, Fritchman JL, Geoghagen NSM, Glodek A, Gnehm CL et al. (1995). Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature **377**(6547):3-17.

Aicardi J (1992). Malformation of the CNS. In Diseases of the nervous system in childhood, Mac Keith, 1st. ed. Paris, pp.150-156.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990). Basic local alignment search tool. J Mol biol **215**:403-410.

Altschul SF, Boguski MS, Gish W and Wootton JC (1994). Issues in searching molecular sequence databases. Nature Genet **6**:119-129.

Antequera F and Bird A (1993). Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci USA **90**:11995-11999.

Antonarakis SE and Aelst LV (1998). Mind the GAP, Rho, Rab and GDI. Nature.Genet **19**:106-108.

American Psychiatric Association (1995). Diagnostic and Statistical manual of mental disorders, 4th revised ed. (DSM-IV-R) (Washington DC:APA).

Baere ED, Fukushima Y, Small K, Udar N, Camp GV, Verhoeven K, Palotie A, Paepe AD and Messiaen L (2000). Identification of BPESC1, a novel gene disrupted by a balanced chromosomal translocation, t(3;4)(q23;p15.2), in a patient with BPES. Genomics **68**:296-304.

Ballabio A (1993). The rise and fall of positional cloning? Nature Genet **3**:277-279.

Bate L and Gardiner M (1999). Genetics of inherited epilepsies. Epileptic Disorders **1**(1):7-19.

Battey J and Clayton DA (1978). The transcription map of mouse mitochondrial DNA. Cell **14**:143-156.

Bedell M, Jenkins N and Copeland N (1996). Good genes in bad neighbourhoods. Nature Genet **11**:229-232.

Bennett M and Reed R (1993). Correspondence between a mammalian spliceosome component and an essential yeast splicing factor. Science **262**:105-108.

Bercovich D, Regev Z, Ratz T, Luder A, Plotsky Y and Gruenbaum Y. Quantitative ratio of primer pairs and annealing temperatures affecting PCR products in duplex amplification. BioTechniques 1999 **27**(4):762-770.

Bickmore WA and Bird AP (1992). Use of restriction endonucleases to detect and isolate genes from mammalian cells. Methods Enzymol **216**:224-245.

Bird AP (1987). CpG islands as gene markers in the vertebrate nucleus. Trends in Genet **3**:342-347.

Boguski MS, Lowe TMJ and Tolstoshev CM (1993). dbEST-datbase for "expressed sequence tags". Nature Genet **4**:332-333.

Boughman JA, Stick MJ, Peterson DA and Cohen MM (1992). Linkage analysis and predicting genetic disease. Clin Lab Med **12**(3):449-61.

Boycott KM, Zahorchak RJ, Summer CG, Boycott NP, Kotak V, Russell CG and Bech-Hansen NT (1998). Construction of a 1.5-Mb bacterial artificial chromosome contig in Xp11.23, a region of high gene content. Genomics **48**:369-372.

Brenner S (1990). The human genome: the nature of the enterprise. CIBA Found Symp **149**:6-17.

Brown T (1993). Analysis of DNA sequences by blotting and hybridisation. In Current protocols in molecular biology, Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA and Struhl K, eds. (New York: John Wiley and Sons, Inc.), pp:2.9.1-2.9.20.

Buckler AJ, Chang DD, Graw SL, Brook JD, Haber DA, Sharp PA and Housman DE (1991). Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. Proc Natl Acad Sci USA **88**:4005-4009.

Buchanan N (1995). The aetiology, epidemiology and prognosis of epilepsy. In Epilepsy a handbook (London: WB Saunders), pp:13-17.

Burn TC, Connors TD, Klinger KW and Landes GM (1995). Increased exon-trapping efficiency through modifications to the pSPL3 splicing vector. Gene **161**:183-187.

Burke DT, Carle GF and Olson MV (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science **236**:806-812.

Callen DF (1986). A mouse-human hybrid cell panel for mapping human chromosome 16. Ann Genet **29**(4):235-239.

Callen DF, Baker E, Eyre HJ Chernos JE, Bell JA and Sutherland GR (1990a). Reassessment of two apparent deletions of chromosome 16p to an ins (11,16) and a t(1;16) by chromosome painting. Ann Genet **33**: 219-221.

Callen DF, Baker E, Eyre HJ, Lane SA (1990b). An expanded mouse-human hybrid cell panel for mapping human chromosome 16. Ann Genet **33**(4):190-195.

Callen DF, Doggett NA, Stallings RL, Chen LZ, Whitmore SA, Lane SA, Nancarrow JK, Apostolou S, Thompson AD, Lapsys NM, Eyre HJ, Baker EG, Shen Y, Holman K, Phillips H, Richards RI and Sutherland GR (1992). High-resolution cytogenetic-based physical map of human chromosome 16. Genomics **13**:1178-1185.

Callen DF, Lane SA, Kozman H, Kremmidiotis G, Whitmore SA, Lowenstein M, Doggett NA, Kenmochi N, Page DC, Maglott DR, Nierman WC, Murakawa K, Berry R, Sikela JM, Houlgatte R, Auffary C and Sutherland GR (1995). Integration of transcript and genetic maps of chromosome 16 at near-1-Mb resolution: demonstration of a "hot spot" for recombination at 16p12. Genomics **29**:503-511.

Cao Y, Kang HL, Xu X, Wang M, Dho SH, Huh JR, Lee BJ, Kalush F, Bocskai D, Ding Y, Tesmer JG, Lee J, Moon E, Jurecic V, Baldini A, Weier H-U, Doggett NA, Simon MI, Adams MD and Kim UJ (1999). A 12-Mb complete coverage BAC contig map in human chromosome 16p13.1-11.2. Genome Res **9:**763-774.

Chaplin DD and Brownstein BH (1995). Yeast artificial chromosome libraries. In Current protocols in molecular biology, Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA and Struhl K, eds (New York: John Wiley and Sons, Inc), pp:6.10.4-6.10.5.

Charlier C, Singh NA, Ryan SG, Lewis TB, Reus BE, Leach RJ and Leppert M (1998). A pore mutation in a novel KQT-like potassium channel gene in an idiopathic epilepsy family. Nature Genet **18**:53-55.

Chelly J (1999). Breakthroughs in molecular and cellular mechanisms underlying X-linked mental retardation. Hum Mol Genet **8**(10):1833-1838.

Chomczynski P and Sacchi N (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem **162**:156-159.

Chu G, Vollrath D and Davis RW (1986). Separation of large DNA molecules by contour-clamped homogeneous electric fields. Science **234**:1582-1585.

Chung CT, Niemela SL and Miller RH (1989). One step preparation of competent Escherichia coli: Transformation and storage of bacterial cells in the same solution. Proc Natl Acad Sci USA **86**:2172-2175.

Chumakov IM, Rigault P, Gall IL, Bellanne-Chantelot C, Billault A, Guillou S, Soularue P, Guasconi G, Poullier E, Gros I, Belova M, Sambucy JL, Susini L, Gervy P, Glibert F, Beaufils S, Bui H, Massart C, De Tand MF, Dukasz F, Lecoulant S, Ougen P, Perrot V, Saumier M, Soravito C et al. (1995). A YAC contig map of the human genome. Nature **377**:175-183.

Claverie JM (1997). Computational methods for the identification of genes in vertebrate genomic sequences. Hum Mol Genet **6**:1735-1744.

Cockell M and Gasser SM (1999). Nuclear compartments and gene regulation. Curr Opin Genet Dev **9**:199-205.

Cohen D, Chumakov I and Weissenbach J (1993). A first-generation physical map of the human genome. Nature **366**(6456):698-701.

Collins FS (1992). Positional cloning lets not call it positional any more. Nature Genet **1**:3-6.

Collins FS (1995). Positional cloning moves from perditional to traditional. Nature Genet **9**:347-350.

Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R and Walters L (1998). New goals for the U.S. Human Genome Project: 1998-2003. Science **282**:682-689.

Cooper DN, Smith BA, Cooke HJ, Niemann S, and Schmidtke J (1985). An estimate of unique DNA sequence heterozygosity in the human genome. Hum Genet **69**:201-205.

Corbo L, Maley JA, Nelson DL and Caskey CT (1990). Direct cloning of human transcripts with hnRNA from hybrid cell lines. Science **249**:652-655.

Crollius HR, Jaillon O, Bernot A, Dasilva C, Bouneau L, Fischer C, Fizames C, Wincker P, Brottier P, Quetier F, Saurin W and Weissenbach J (2000). Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. Nature Genet **25**:235-238.

de Jong PJ, Yokabata K, Chen C, Lohman F, Pedersen L, McNinch J and Van Dilla M (1989). Human chromosome-specific partial digest libraries in *k* and cosmid vectors. Cytogenet Cell Genet **58**:985.

Delgado-Escueta AV, Serratosa JM, Liu A, Weissbecker K, Medina MT, Gee M, Trieman LJ and Sparkes RS (1994). Progress in Mapping Human Epilepsy Genes. Epilepsia **35**:S29-S40.

Deloukas P, Schuler GD, Gyapay G, Beasley EM, Soderlund C, Rodriguez-Tome P, Hui L, Matise TC, McKusick KB, Beckmann JS, Bentolila S, Bihoreau MT, Birren BB, Browne J, Butler A, Castle AB, Chiannilkulchai N, Clee C, Day PJR, Dehejia A, Dibling T, Drouot N, Duprat S, Fizames C, Fox S et al. (1998). A physical map of 30,000 human genes. Science **282**: 744-746.

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J and Weissenbach J (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature **380**:152-154.

Doggett NA, Stallings RL, Hildebrand CE and Moyzis RK (1992). The mapping of Chromosome 16. In the Human Genome Project, Los Alamos National laboratory. Los Alamos Science **20**:182-209.

Doggett NA, Goodwin LA, Tesmer JG, Meincke LJ, Bruce DC, Clark LM, Altherr MR, Ford AA, Chi HC, Marrone BL, Longmire JL, Lane SA, Whitmore SA, Lowenstein MG, Sutherland RD, Mundt MO, Knill EH, Bruno WJ, Macken CA, Torney DC, Wu JR, Griffith J, Sutherland GR, Deaven LL, Callen DF and Moyzis RK (1995). An integrated physical map of human chromosome 16. Nature **377**(28):335-339.

Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, Lander ES, Botstein D, Akots G, Rediker KS, Gravius T, Brown VA, Rising MB, Parker C, Powers JA, Watt DE, Kauffman ER, Bricker A, Phipps P, Muller-Kahle H, Fulton TR, Siu NG et al. (1987). A genetic linkage map of the human genome. Cell **51**:319-337.

Dunham I, Hunt AR, Collins JE, Bruskiewich R, Beare DM, Clamp M, Smink LJ, Ainscough R, Almeida JP, Babbage A, Bagguley C, Bailey J, Barlow K, Bates KN, Beasley O, Bird CP, Blakey S, Bridgeman AM, Buck D, Burgess J, Burrill WD, Burton J, Shimizu N, Roe BA, Chissoe S et al. (1999). The DNA sequence of human chromosome 22. Nature **402**:489-495.

Elmslie FV, Rees M, Williamson MP, Kerr M, Kjeldsen MJ, Pang KA, Sundqvist A, Friis ML, Chadwick D, Richens A, Covanis A, Santos M, Arzimanoglou A, Panayiotopoulos CP, Curtis D, Whitehouse WP and Gardiner M (1997). Genetic mapping of a major susceptibility locus for juvenile myoclinc epilepsy on chromosome 15q. Hum Mol Genet **6**:1326-1334.

Elvin P, Slynn G, Black D, Graham A, Butler R, Riley J, Anand R and Markham AF (1990). Isolation of cDNA clones using yeast artificial chromosome probes. Nucleic Acids Res **18**:3913-3917.

Escayg A, MacDonald BT, Meisler MH, Baulac S, Huberfeld G, An-Gourfinkel I, Brice A, LeGuern E, Moulard B, Chaigne D, Buresi C and Malafosse A (2000). Mutations of SCN1A, encoding a neuronal sodium channel, in two families with GEFS+2. Nature Genet **24**(4):343-345.

Ewing B and Green P (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res **8**(3):186-194.

Ewing B and Green P (2000). Analysis of expressed sequence tags indicates 35,000 human genes. Nature Genet **25**:232-234.

Fearon ER, Cho KR, Nigro JM, Kern SE, Simons JW, Ruppert JM, Hamilton SR, Preisinger AC, Thomas G, Kinzler KW and Vogelstein B (1990). Identification of a chromosome 18q gene that is altered in colorectal cancers. Science **247**:49-56.

Fletcher CF and Frankel WN (1999). Ataxic mouse mutants and molecular mechanisms of absence epilepsy. Hum Mol Genet **8**(10):1907-1912.

Fields C, Adams MD, White O and Venter JC (1994). How many genes in the human genome? Nature Genet **7**:345-346.

Finney M (1994). Pulsed-field gel elctrophoresis. In Current protocols in molecular biology, Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA and Struhl K, eds. (New York: John Wiley & Sons, Inc.), pp:2.5.9-2.5.17.

Gardiner-Garden M and Frommer M (1987). CpG islands in vertebrate genomes, J Mol Biol **196**:261-282.

Gecz J, Gedeon AK, Sutherland GR and Mulley JC (1996). Identification of the gene FMR2, associated with FRAXE mental retardation. Nature Genet **13**(1):105-108.

Gecz J, Bielby S, Sutherland GR and Mulley JC (1997). Gene structure and subcellular localization of FMR2, a member of a new family of putative transcription activators. Genomics **44**:201-213.

Gecz J (2000a). The FMR2 gene, FRAXE and non-specific X-linked mental retardation: clinical and molecular aspects. Ann Hum Genet **64**:95-106.

Gecz J (2000b). FMR3 is a novel gene associated with FRAXE CpG island and transcriptionally silent in FRAXE full mutations. J Med Genet **37**:782-784.

Gecz J, Barnett S, Liu J, Hollway G, Donnelly A, Eyre H, Eshkevari HS, Baltazar R, Grunn A, Nagaraja R et al., (2000). Characterization of the human glutamate receptor subunit 3 gene (GRIA3), a candidate for bipolar disorder and non-specific X-linked mental retardation. Genomics (in press).

Gecz J and Mulley J (2000). Genes for cognitive function: developments on the X. Genome Res **10**(2):157-63.

Gedeon AK, Donnelly AJ, Mulley JC, Kerr B and Turner G (1996). How many X-linked genes for non-specific mental retardation (MRX) are there? [letter]. Am J med Genet **64**:158-162.

Geraghty MT, Brody LC, Martin LS, Marble M, Kearns W, Pearson P, Monaco AP, Lehrach H and Valle D (1993). The isolation of cDNAs from OATL1 at Xp11.2 using a 480-kb YAC. Genomics **16**:440-446.

Giles RH, Petrij F, Dauwerse HG, Den Hollander AI, Lushnikova T, Van Ommen GJB, Goodman RH, Deaven LL, Doggett NA, Peters DJM and Breuning MH (1997). Construction of a 1.2-Mb contig surrounding, and molecular analysis of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome 16p13.3. Genoimcs **42**:96-114.

Glass IA (1991). X linked mental retardation. J Med Genet **28**:361-371.

Goss SJ and Harris H (1975). New method for mapping genes in human chromosomes. Nature **255**:680-684.

Greenberg DA, Delgado-Escueta AV, Widelitz H, Sparkes RS, Treiman LJ, Maldonado HM and Terasaki PI (1988). Juvenile myoclonic epilepsy (JME) may be linked to the Bf and HLA loci in human chromosome 6. Am J Med Genet **31**:185-192.

Grunstein M and Hogness DS (1975). Colony hybridisation: A method for the isolation of cloned DNAs that contain a specific gene. Proc Natl Acad Sci USA **72**:3961-3965.

Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'Homme JF, Dib C, Auffery C, Morissette J, Weissenbach J and Goodfellow PN (1996). A radiation hybrid map of the human genome. Hum Mol Genet **5**(3):339-346.

Habuchi T, Luscombe M, Elder PA and Knowles A (1998). Structure and methylation-based silencing of a gene (DBCCRI) within a candidate bladder cancer tumor suppressor region at 9q32-q33. Genomics **48**:277-288.

Hagberg B and Kyllerman M (1983). Epidemiology of mental retardation. A Sewidish survey. Brain Dev **5**:441-449.

Hattori H, Hayashi K, Okuno T, Temma S, Fujii T, Ochi J and Mikawa H (1985). De novo reciprocal translocation t(6;14)(q27;q13.3) in a child with infantile spasms. Epilepsia **26**(4):310-313.

Hattori M, Fujiyama A, Taylor TD, Watanabe H, Yada T, Park HS, Toyoda A, Ishii K, Totoki Y, Choi DK, Soeda E, Okhi M, Takagi T, Sakaki Y, Taudien S, Blechschmidt K, Polley A, Menzel U, Delabar J, Kumpf K, Lehmann R, Patterson D, Reichwald K, Rump A, Schillhabel M et al. (2000). The DNA sequence of human chromosome 21. Nature **405**:311-319.

Hauser WA, Annegers JF and Rocca WA (1996). Descriptive epidemiology of epilepsy: contributions of population-based studies from Rochester, Minnesota. Mayo Clin Proc **71**:576-586.

Heller RA, Schena M, Chai A, Shalon D, Bedilion T, Gilmore J, Woolley DE and Davis RW (1997). Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. Proc Natl Acad Sci USA **94**(6):2150-2155.

Hildebrand CE, Torney DC and Wagner RP (1992). Informativeness of polymorphic DNAmarkers. In the Human Genome Project, Los Alamos National laboratory. Los Alamos Science **20**:100-102.

Higgins JJ, Rosen DR, Loveless JM, Clyman JC and Grau MJ (2000). A gene for nonsyndromic mental retardation maps to chromosome 3p25-pter. Neurology **55**(3):335-40.

Hofmann K, Bucher P, Falquet L and Bairoch A (1999). The PROSITE database, its status in 1999. Nucleic Acids Res **27**(1):215-219.

Holinski-Feder E, Reyniers E, Uhrig S, Golla A, Wauters J, Kroisel P, Bossuyt P, Rost I, Jedele K, Zierler H, Schwab S, Wildenauer D, Speicher MR, Willems PJ, Meitinger T and Kooy RF (2000). Familial mental retardation syndrome ATR-16 due to an inherited cryptic subtelomeric translocation, t(3;16)(q29;p13.3). Am J Hum Genet **66**(1):16-25.

Howell MW, Jobs M, Gyllensten ULF and Brookes AJ (1999). Dynamic allele-specific hybridisation. A new method for scoring single nucleotide polymorphisms. Nature Biotech **17**:87-88.

Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu SH, 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 et al. (1995). An STS-based map of the human genome. Science **270**:1945-1954.

Huynh DP, Del Bigio MR, Ho DH and Pulst SM (1999). Expression of ataxin-2 in brains from normal individuals and patients with Alzheimer's disease and spinocerebellar ataxia 2. Ann Neurol **45**(2):232-241.

Johnson EW, Dubovsky J, Rich SS, O'Donovan CA, Orr HT, Anderson VE, Gil-Nagel A, Ahmann P, Dokken CG, Schneider DT and Weber JL (1998). Evidence for a novel gene for familial febrile convulsions, FEB2, linked to chromosome 19p in an extended family from the Midwest. Hum Mol Genet **7**(1):63-67.

Karpen GH (1994). Position-effect variegation and the new biology of heterochromatin. Curr Opin Genet Dev **4**:281-291.

Kato R and Sasaki H (1998). Quick identification and localization of CpG islands in large genomic fragments by partial digestion with *Hpa* II and *Hha* I. DNA Research **5**:287-295.

Kerr B, Turner G, Mulley J, Gedeon A and Partington M (1991). Nonspecific X linked mental retardation. J Med Genet **28**:378-382.

Kleinjan DJ, and Heyningen VV (1998). Position effect in human genetic disease. Hum Mol Genet **7**(10):1611-1618.

Knight SJL, Flannery AV, Hirst MC, Campbell L, Christodoulou Z, Phelps SR, Pointon J, Middleton-Price HR, Barnicoat A, Pembrey ME, Holland J, Oostra BA, Bobrow M and Davies KE (1993). Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell **74**:127-143.

Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C and Kunkel LM (1987). Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell **50**:509-517.

Kogan SC, Doherty BS and Gitsher J (1987). An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. N Eng J Med **317**:985-990.

Korn B, Sedlacek Z, Manca A, Kioschis P, Konecki D, Lehrach H and Poustka A (1992). A strategy for the selection of transcribed sequences in the Xq28 region. Hum Mol Genet **1**:235-242.

Kozak M (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translocation. J Biol Chem **226**(30):19867-19870.

Kozak M (1996). Interpreting cDNA sequences: some insights from studies on translation. Mamm Genome **7**:563-574.

Kozman HM, Keith TP, Donis-Keller H, White RL, Weissenbach J, Dean M, Vergnaud G, Kidd K, Gusella J, Royle NJ, Sutherland GR and Mulley JC (1995). The CEPH consortium linkage map of human chromosome 16. Genomics **25**:44-58.

Landegren U (1992). Detection of mutations in human DNA. Genet Anal Tech Appl **9**(1):3-8.

Larsen F, Gundersen G, Lopez R and Prydz H (1992). CpG islands as gene markers in the human genome. Genomics **13**:1095-1107.

Ledbetter DH, Rich DC, O'Connell P, Leppert M and Carey JC (1989). Precise localisation of NF1 to 17q11.2 by balanced translocation Am J Hum Genet **44**:20-24.

Leshima A and Takeshita K (1998). Chromosome abnormalities and epileptic seizures. Jpn J Hum Genet **33**:49-60.

Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL and Quackenbush J (2000). Gene index analysis of the human genome estimates approximately 120,000 genes. Nature Genet **25**:239-240.

Liu P, Legerski R and Siciliano MJ (1989). Isolation of human transcribed sequences from human-rodent somatic cell hybrids. Science **246**:813-815.

Liu P, Perryman MB, Liao W and Siciliano MJ (1992). Efficiency and limitations of the hn-cDNA library approach for the isolation of human transcribed genes from hybrid cells. Somat Cell Mol Genet **18**:7-18.

Lopes-Cendes I, Scheffer IE, Berkovic SF, Rousseau M, Andermann E and Rouleau GA (2000). Anew locus for generalized epilepsy with febrile seizures plus maps to chromosome 2. Am J Hum Genet **66**:698-710.

Lubs HA, Chiurazzi P, Arena JF, Schwartz C, Tranebjaerg L and Neri G (1996). XLMR genes: update 1996. Am J Med Genet **64**:147-157.

Lunetta KL, Boehnke M, Lange K and Cox DR (1996). Selected locus and multiple panel models for radiation hybrid mapping. Am J Hum Genet **59**:717-725.

Maarek O, Salabelle A, Coniat MBL, Daniel MT and Berger R (1999). Chromosome 16 inversion-associated translocation: two new cases. Cancer Genet Cytogenet **114**:126-129.

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

Marquardt A, Stöhr H, Passmore LA, Krämer F Rivera A and Weber BHF (1998). Mutations in a novel gene, VMD2, encoding a protein of unknown properties cause juvenile-onset vitelliform macular dystrophy (Best's disease). Hum Mol Genet **7**(9):1517-1525.

McLaren J and Bryson SE (1987). Review of recent epidemiological studies of mental retardation: Prevalence, associated disorders, and etiology. Am J Ment Retard **92**:243-254.

Mermod N, O'Neill EA, Kelly TJ and Tjian R (1989). The proline-rich transcriptional activator of CTF/NF-1 is distinct from the replication and DNA binding domain. Cell **58**:741-753.

Milot E, Fraser P and Grosveld F (1996). Position effects and genetic disease. TIG **12**(4):123-126.

Minassian BA, Lee JR, Herbrick JA, Huizenga J, Soder S, Mungall AJ, Dunham I, Gardner R, Fong CYG, Carpenter S, Jardim L, Satishchandra P, Andermann E, Carter Snead III O, Lopes-Cendes I, Tsui LC, Delgado-Escueta AV, Rouleau GA and Scherer SW (1998). Mutations in a gene encoding a novel protein tyrosine phosphatase cause progressive myoclonus epilepsy. Nature Genet **20**:171-174.

Monaco AP, Neve RL, Colletti-Feener C, Bertelson CJ, Kurnit DM and Kunkel LM (1986). Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. Nature **323**:646-650.

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

Morton NE (1978). Effect of inbreeding on IQ and mental retardation. Proc Natl Acad Sci **75**:3906-3908.

Morton NE (1991). Parameters of the human genome. Proc Natl Acad Sci USA **88**:7474-7476.

Moore D (1993). Preparation and analysis of DNA. In Current protocols in molecular biology, Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA and Struhl K, eds. (New York: John Wiley and Sons, Inc.), pp:2.0.5-2.1.9.

Mulley JC and Sutherland GR (1993). Integrating maps of chromosome 16. Curr.Op.Genet.Dev **3**:425-431.

Munroe DJ, Haas M, Bric E, Whitton T, Aburatani H, Hunter K, Ward D and Housman DE (1994). IRE-bubble PCR: A rapid method for efficient and representative amplification of human genomic DNA sequences from complex sources. Genomics **19**:506-514.

Nakai K and Horton P (1999). PSORT: a program for detecting sorting signls in proteins and predicting their subcellular localization. Trends Biochem Sci **24**(1):34-36.

Nigro CL, Faravelli F, Cavani S, Perroni L, Novello P, Vitali M, Bricarelli FD and Grasso M (2000). FRAXE mutation in a mentally retarded subject and his phenotypically normal twin brother. Eur J Hum Genet **8**:157-162.

Noebels JL (1996). Targeting epilepsy genes. Neuron **16**:241-244.

Olson M, Dutchik JE, Graham MY, Brodeur GM, Helms C, Frank M, MacCollin M, Scheinman R and Frank T (1986). Random-clone strategy for genomic restriction mapping in yeast. Proc Natl Acad Sci USA **83**:7826-7830.

Olson M, Hood L, Cantor C and Botstein D (1989). A common language for physical mapping of the human genome. Science **245**:1434-1435.

Orkin SH (1986). Reverse genetics and human disease. Cell **47**:845-850.

Parimoo S, Patanjali SR, Shukla H, Chaplin DD and Weissman SM (1991). CDNA selection: Efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. Proc Natl Acad Sci USA **88**:9623-9627.

Peiffer A, Thompson J, Charlier C, Otterud B, Varil T, Pappas C, Barnitz C, Gruenthal K, Kuhn R and Leppert M (1999). A locus for febrile seizures (FEB3) maps to chromosome 2q23-24. Ann Neurol **46**(4):671-678.

Phillips HA, Scheffer IE, Berkovic SF, Hollway GE, Sutherland GR, and Mulley JC (1995). Localization of a gene for autosomal dominant nocturnal frontal lobe epilepsy to chromosome 20q13.2. Nature Genet **10**(1):117-118.

Phillips HA, Marini C, Scheffer IE, Sutherland GR, Mulley JC and Berkovic SF (2000). A de novo mutation in sporadic frontal lobe epilepsy. Ann Neurol **48**(2):264-267.

Phillips HA, Favre I, Kirkpatrick M, Zuberi SM, Goudie D, Heron SE, Scheffer IE, Sutherland GR, Berkovic SF, Bertrand D and Mulley JC (2000). CHRNB2 is the second acetylycholine receptor subunit associated with autosomal dominant nocturnal frontal lobe epilepsy. Am J Hum Genet **68**:(in press).

Pulst SM (1999). Genetic linkage analysis. Arch Neurol **56**(6):667-672.

Reed KC and Mann DA (1985). Rapid transfer of DNA from agarose gels to nylon membranes. Nucl Acids Res **13**:7207-7221.

Reddy KS, Rajangam S and Thomas IM (1999). Structural chromosomal anomaly in mental retardation. Indian J Pediatr **66**(6):937-940.

Rimoin DL, Connor MJ and Pyeritz RE (1997). Nature and frequency of genetic disease. In Emery and Rimoin's principles and practise of medical genetics, 3rd. eds. (New York: Churchill Livingstone), Vol 1 pp:31-34.

Robinson R and Gardiner M (2000). Genetics of childhood epilepsy. Arch Dis Child **82**:121-125.

Rommens JM, Iannuzzi MC, Kerem BS, Drumm MJ, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N, Zsiga M Buchwald M, Riordan JR, Tsui LC and Collins FC (1989). Identification of Cystic fibrosis gene: chromosome walking and jumping. Science **245**:1059-1065.

Rommens JM, Lin B, Hutchinson GB, Andrew SE, Goldberg YP, Glaves ML, Graham R, Lai V, McArthur J, Nasir J, Theilmann J, McDonald H, Kalchman M, Clarke LA, Schappert K and Hayden MR (1993). A transcription map of the region containing the Huntington disease gene. Hum Mol Genet **2**(7):901-907.

Rosenberg SAM, Knight S, Gaunt L, Fergusson W, Killoran C, Smith JC, Kingston H, Campbell RH, Flint J, Donnai D and Beisecker L (1999). Screening for submicroscopic chromosome rearrangements in children with idiopathic mental retardation using microsatellite markers for the chromosome telomeres. J Med Genet **36**:405-411.

Royer-Pokora B, Kunkel LM, Monaco AP, Goff SC, Newburger PE, Baehner RL, Cole SF, Curnutte JT and Orkin SH (1986). Cloning the gene for an inherited human disorder-chronic granulomatous disease-on the basis of its chromosomal location. Nature **322**:32-38.

Ruddle FH (1984). The William Allan memorial award address reverse genetics and beyond. Am J Hum Genet **36**:944-953.

Ryan SG (1999). Ion channels and the genetic contribution to epilepsy. J Child Neurol **14**:58-66.

Sander T, Schulz H, Saar K, Gennaro E, Riggio C, Bianchi A, Zara F, Luna D, Bulteau C, Kaminska A, Ville D, Cieuta C, Picard F, Prud'homme JF, Bate L, Sundquist A, Gardiner RM, Janssen GAMAJ, Jan de haan G, Dorothee GA, Trenite KN, Bader A, Lindhout D, Riess O, Wienker TF, Janz D and Reis A (2000). Genome search for susceptibility loci of common idiopathic generalised epilepsies. Hum Mol Genet **9**(10):1465-1472.

Sambrook J, Fritsch EF and Maniatis T (1989). Molecular cloning: A laboratory mannual, Vol 1 (New York: Cold Spring Harbor Laboratory Press).

Scheffer IE, Wallace RH, Mulley JC and Berkovic SF (2000). Locus for febrile seizures. Ann Neurol **47**:840-841.

Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE, Tome PR, Aggarwal A, Bajorek E, Bentolila, S, Birren BB, Butler A, Castle AB,
Chiannilkulchai N, Chu A, Clee C, Cowles S, Day PJR, Dibling T, Drouot N, Dunham I, Duprat S, East C, Edwards C, Fan JB, Fang N, Fizames C, Garrett C, Green L, Hadley D et al. (1996). A gene map of the human genome. Science **274**:540-545.

Schwartz DC, Saffron W, Welsh J, Haas R, Goldenberg M and Cantor CR (1982). New techniques for purifying large DNAs and studying their properties and packaging. Cold Spring Harbor Symp. Quant Biol **47**:189-195.

Schwarz H and Whitton JL (1992). A rapid, inexpensive method for eluting DNA from agarose or acrylamide gel slices without using toxic or chaotropic materials. BioTechniques **13**(2):205-206.

Schweizer D (1981). Counterstain enhanced chromosome banding. Hum Genet **57**:1-14.

Sealy PG, Whittaker PA and Southern EM (1985). Removal of repeated sequences from hybridisation probes. Nuc Acids Res **13**:1905-1922.

Serratosa JM, Gardiner RM, Lehesjoki AE, Pennacchio LA and Myers RM (1999). The molecular genetic bases of the progressive myoclonus epilepsies. Adv Neurol **79**:383-398.

Serratosa JM, Gomez-Garre P, Gallardo ME, Anta B, Beltran-Valero de Bernabe D, Lindhout D, Augustijn PB, Tassinari CA, Michelucci R, Malafosse A, Topcu
M, Grid D, Dravet C, Berkovic SF and Rodriguez de Cordoba S (1999). A novel protein tyrosine phosphatase gene is mutated in progressive myoclonus epilepsy of the Lafora type (EPM2). Hum Mol Genet **8**(2):345-352.

Shapiro MB and Senapathy P (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res **15**(17):7155-7174.

Sharp PA, Berk AJ and Berget SA (1980). Transcription maps of adenovirus. Methods Enzymol **65**:750-768.

Sheets MD, Ogg SC and Wickens MP (1990). Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. Nucleic Acids Res **18**(19):5799-5805.

Shen Y, Kozman HM, Thompson A, Phillips HA, Holman K, Nancarrow J, Lane S, Chen LZ, Apostolou S, Doggett NA, Callen DF, Mulley JC, Sutherland GR and Richards RI (1994). A PCR-based genetic linkage map of human chromosome. Genomics **22**:68-76.

Shepherd NS, Pfrogner BD, Coulby JN, Ackerman SL, Vaidyanathan G, Sauer RH, Balkenhol TC and Sternberg N (1994). Preparation and Screening of an arrayed human genomic library generated with the P1 cloning system. Proc Natl Acad Sci USA **91**:2629-2633.

Shibata H, Huynh DP and Pulst SM (2000). A novel protein with RNA-binding motifs interacts with ataxin-2. Hum Mol Genet **9**(9):1303-1313.

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y and Simon M (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc Natl Acad Sci USA **89**:8794-8797.

Singh NA, Charlier C, Stauffer D, DuPont BR, Leach RJ, Melis R, Ronen GM, Bjerre I, Quattlebaum T, Murphy JV, McHarg ML, Gagnon D, Rosales TO, Peiffer A, Anderson VE and Leppert M (1998). A novel potassium channel gene, KCNQ2, is mutated in an inherited epilepsy of newborns. Nature Genet **18**:25-29.

Southern EM, (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol **98**:503-507.

St Clair D, Blackwood D, Muir W, Carothers A, Walker M, Spowart G, Gosden C and Evans HJ (1990). Association within a family of a balanced autosomal translocation with major mental illness. Lancet **336**(8706):13-16.

Steffenburg U, Hagberg G, Viggedal G and Kyllerman M (1995). Active epilepsy in mentally retarded children. I. Prevalence and additional neuro-impairments. Acta Paediatr **84**(10):1147-1152.

Steinlein OK, Mulley JC, Propping P, Wallace RH, Phillips HA, Sutherland GR, Scheffer IE and Berkovic SF (1995). A missense mutation in the neuronal nicotinic acetylcholine receptor α4 subunit is associated with autosomal dominant nocturnal frontal lobe epilepsy. Nature Genet **11**:201-203.

Steinlein OK (1998). New insights into the molecular and genetic mechanisms underlying idiopathic epilepsies. Clin Genet **54**:169-175.

Sternberg N (1990). Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. Proc Natl Acad Sci USA 87:103-107.

Stevenson R (1996). Mental retardation: overview and historical perspective. Proc Greenwood Genet Ctr **15**:19-25.

Stewart EA, McKusick KB, Aggarwal A, Bajorek E, Brady S, Chu A, Fang N, Hadley D, Harris M, Hussain S, Lee R, Maratukulam A, Connor KO, Perkins S, Piercy M, Qin F, Reif T, Sanders C, She X, Sun WL, Tabar P, Voyticky S, Cowles S, Fan JB, Mader C et al. (1997). An STS-based radiation hybrid map of the human genome. Genome Res **7**:422-433.

Stromme P and Hagberg G (2000). Aetiology in severe and mild mental retardation: a population-based study of Norwegian children. Developmental Med & Child Neurol **42**:1-10.

Sutherland GR and Baker E (1992). Characterization of a new rare fragile site easily confused with fragile X. Hum Mol Genet **1**:111-113.

Tempkin O (1971). The falling sickness. In A history of epilepsy from the Greeks to beginning of modern neurology, 2nd edition (Baltimore and London: John Hopkins Press).

Tharapel AT, Summitt RL, Wilroy RS Jr. and Martens P (1977). Apparently balanced *de novo* translocations in patients with abnormal phenotypes: Report of 6 cases. Clin Genet **11**:255-269.

Toniolo D and D'Adamo P (2000). X-linked non-specific mental retardation. Curr Opin Genet Dev **10**:280-285.

Tsuji S, Choudary PV, Martin BM, Stubblefield BK, Mayor JA, Barranger JA and Ginns EI (1987). A mutation in the human glucocerebrosidase gene in neuronopathic Gaucher's disease. N Eng J Med **316**(10):570-575.

Turner G, Webb S and Robinson H (1996). Prevalence of fragile X syndrome. Am J Med Genet **64**:158-162.

Van der Maarel SM, Scholten IH, Huber I, Philippe C, Suijkerbuijk RF, Gilgenkrantz S, Kere J, Cremers FP and Ropers HH (1996). Cloning and characterization of DXS6673E, a candidate gene for X-linked mental retardation in Xq13.1. Hum Mol Genet. **5**:887-897.

Verkerk AJMH, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DPA, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang F, Eussen BE, Van Ommen GJB, Blonden LAJ, Riggins GJ, Chastain JL, Kunst CB, Galjaard H, Caskey CT, Nelson DL, Oostra BA and Warren ST (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome Cell **65**:905-914.

Villard L, Briault S, Lossi AM, Paringaux C, Belounge J, Colleaux L, Pincus DR, Woollatt E, Lespinasse J, Munnich A, Moraine C, Fontes M and Gecz J (1999). Two unrelated patients with inversion of the X chromosome and non-specific mental retardation: physical and transcriptional mapping of their common breakpoint region in Xq13.1. J Med Genet **36**:754-758.

Voltz AK, Braun A, Chines PS, Doheny KF, Hilliard MS, Little DP, Witmer PD and Nussbaum RL (2000). Developing a SNP map of the human genome. Am J Hum Genet **67**(4) Supplement 2:336.

Wahle E and Keller W (1992). The biochemistry of 3'-end cleavage and ployadenylation of messenger RNA precursors. Annu Rev Biochem **61**:419-440.

Wallace RH, Berkovic SF, Howell RA, Sutherland GR and Mulley JC (1996). Suggestion of a major gene for familial febrile convulsions mapping to 8q13-21. J Med Genet **33**:308-312.

Wallace RH, Wang DW, Singh R, Scheffer IE, George AL, Jr, Phillips HA, Saar K, Reis A, Johnson EW, Sutherland GR, Berkovic SF and Mulley JC (1998). Febrile seizures and generalized epilepsy associated with a mutation in the Na^+-channel β1 subunit gene SCN1B. Nature Genet **19**:366-370.

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M and Lander ES (1998). Large- scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science **280**:1077-1082.

Warburton D (1991). De novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. Am J Hum Genet **49**(5):995-1013.

Weber JL and May PE (1989). Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. Am J Hum Genet **44**:388-396.

Weissenbach J (1993). A second generation linkage map of the human genome based on highly informative microsatellite loci. Gene **135**:275-278.

Whitmore SA, Apostolou S, Lane S, Nancarrow JK, Phillips HA, Richards RI, Sutherland GR and Callen DF (1994). Isolation and characterization of transcribed sequences from a chromosome 16 hn-cDNA library and the physical mapping of genes and transcribed sequences using a high-resolution somatic cell panel of human chromosome 16. Genomics **20**:169-175.

Wilbur WJ and Lipman DJ (1983). Improved tools for biological sequence comparision. Proc Natl Acad Sci **80**:726-730.

Wilcox AS, Khan AS, Hopkins, JA and Sikela JM (1991). Use of 3' untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implications for an expression map of the genome. Nucleic Acids Res **19**:1837-1843.

Zara F, Bianchi A, Avanzini G, Donato SD, Castellotti B, Patel PI, and Pandolfo M (1995). Mapping of genes predisposing to idiopathic generalized epilepsy.Hum Mol Genet **4**(7):1201-1207.

Zara F,Gennaro E, Stabile M, Carbone I, Malacarne M, Majello L, Santangelo R, Antonio de Falco F and Bricarelli FD (2000). Mapping of a locus for a familial autosomal recessive idiopathic myoclonic epilepsy of infancy to chromosome 16p13. Am J Hum Genet **66**:1552-1557.

Zuberi SM, Eunson LH, Spauschus A, De Silva R, Tolmie J, Wood NW, McWilliam RC, Stephenson JPB, Kullmann DM and Hanna MG (1999). A novel mutation in the human voltage-gated potassium channel gene (*Kv1.1*) associates with episodic ataxia type 1 and sometimes with partial epilepsy. Brain **122**(5):817-825.

Appendix

Bhalla, K., Eyre, H.J., Whitmore, S.A., Sutherland, G.R. & Callen, D.F. (1999) C16orf5 a novel proline-rich gene at 16p13.3, is highly expressed in the brain.
Journal of Human Genetics, v. 44(6), pp. 383-387