Rearrangement breakpoints and viral integration sites in the human genome

Deborah Burford

Darwin College

October 2007

This dissertation is submitted for the degree of Doctor of Philosophy

Deborah Burford

Rearrangement breakpoints and viral integration sites in the human genome.

Abstract

It is estimated that 1 in 2000 live births have an apparently balanced translocation and that the risk of congenital abnormalities in these carriers is approximately double that seen amongst individuals with normal karyotypes. To date, less than 50 de novo reciprocal balanced translocations have been sequenced across the breakpoint junctions and little is understood about the mechanisms that may cause these non-recurrent rearrangements. The aim of this thesis is to investigate methods of translocation breakpoint mapping. The results showed that array painting onto a Whole Genome Tile Path array followed by array painting onto a custom-made NimbleGen oligonucleotide array and subsequent amplification of the translocation junctions by long range PCR was the most efficient method tested for generating sequence across translocations. Analysis of the sequence generated showed that in the three patients studied, three of the six breakpoints directly disrupted a gene; CENTG2 in a patient with a t(2;7)(q37.3;p15.1) and PTPRZ1 and DACH1 in a patient with a t(7;13)(q31.3;q21.3). As an alternative approach, generation and screening of a custom-made fosmid library using flow sorted derivative chromosomes from an additional patient was used to generate sequence across the translocation breakpoints. This approach is useful for translocations in which additional complexity at the breakpoints confounds other methods for breakpoint mapping and sequencing. The fosmid library approach developed for the mapping of translocation breakpoints has also been applied to the mapping of viral integration into the human genome furthering our understanding of the integrated virus.

Table of contents

1 **Introduction** 1
1.1 The discovery of human chromosomes 1
1.2 Chromosome rearrangements 1
1.2.1 Chromosome translocations 1
1.2.1.1 Frequency of translocations 3
1.2.1.2 Associated phenotypic risk 6
1.2.2 Deletions, duplications, inversions 8
1.2.2.1 Phenotypic effect of deletions, duplications and inversions 10
1.2.2.2 Genomic disorders 10
1.2.3 Analysis of chromosome rearrangements 11
1.2.3.1 Metaphase chromosome banding analysis 12
1.2.3.2 FISH 12
1.2.3.3 Fluorescence activated chromosome sorting 13
1.2.3.4 Somatic cell hybrids 14
1.2.3.5 Southern Blotting 14
1.2.3.6 Comparative genomic hybridisation 14
1.2.3.7 Array painting 18
1.2.3.8 Custom-made libraries 20
1.2.3.9 Sequence analysis 21
1.2.4 Public databases of constitutional rearrangements 22
1.2.5 Normal variation within the human genome 22
1.3 Mechanisms behind chromosomal rearrangements 24
1.3.1 Double strand breaks 24
1.3.2 Homologous recombination 25
1.3.3 Non-homologous end joining 28
1.3.4 Genome architecture associated with rearrangements 29
1.3.4.1 Chi sequences 29
1.3.4.2 Translin associated motifs 29
1.3.4.3 Topoisomerase I and II sites 29
1.3.4.4 Immunoglobulin heptamers 30
1.3.4.5 Mini satellite sequences 30
1.3.4.6 Purine/pyrimidine tracts 30
1.3.4.7 SINE elements 31
1.3.4.8 LINE elements 31
1.3.4.9 Long terminal repeats 31
1.3.4.10 Class II transposons 32
1.3.4.11 Palindromes 32
1.3.4.12 Segmental duplications 32
1.4 Viral Integration into the human genome 34
1.4.1 The Human Herpes Virus 34
1.4.2 Evidence of integration and inheritance 35
1.4.3 The structure of the HHV-6B Virus 36
1.5 Aims of this thesis 37

1.5.1 Use and adaptation of current techniques for the mapping of rearrangement breakpoints in three patients with abnormal phenotypes 37
1.5.2 Development of techniques for the investigation of rearrangement breakpoints 37
1.5.3 Bioinformatic analysis of the genomic architecture surrounding genomic rearrangement breakpoints 38
1.5.4 Application of techniques developed for the mapping of rearrangement breakpoints to the investigation of viral integration in the human genome 38

2 **Materials and methods** 40
2.1 Cell culture 40
2.2 Isolation of Patient DNA 40
2.2.1 Genomic DNA extraction 40
2.2.2 Flow sorting of derivative chromosomes 41
2.3 Amplification of template DNA 42
2.3.1.1 REPLI-g 42
2.3.1.2 DOP PCR 43
2.4 Construction of microarrays 44
2.4.1 Genomic clone microarrays (fosmid and plasmid microarrays) 44
2.4.1.1 Clone selection 44
2.4.1.2 Picking 44
2.4.1.3 DNA prepping 44
2.4.1.4 DOP PCR amplification of plasmid, fosmid and BAC clones 45
2.4.1.5 Amino-linking of DOP PCR products 46
2.4.1.6 Spotting onto slides 47
2.4.2 PCR product arrays 48
2.4.2.1 PCR primer design 48
2.4.2.2 PCR amplification of products for microarray production 48
2.5 Hybridisation and analysis of microarrays 48
2.5.1 Hybridisation and analysis of in-house arrays (PCR, fosmid, WGTP) 48
2.5.1.1 Labelling 49
2.5.1.2 Precipitation of samples 49
2.5.1.3 Hybridisation and washing 50
2.5.1.4 Scanning and image quantification 50
2.5.2 Hybridisation and analysis of NimbleGen oligonucleotide microarrays (carried out by NimbleGen Systems) 51
2.6 Fluorescence in situ hybridisation 51
2.6.1 Preparation of slides for FISH hybridisation 51
2.6.1.1 Metaphase spreads 51
2.6.1.2 Extended chromatin fibres 52
2.6.2 Preparation of clone insert DNA 53
2.6.2.1 Preparation of fosmid clone insert DNA 53

2.6.2.2 Preparation of plasmid clone insert DNA 54
2.6.3 Fluorescence labelling of probes 55
2.6.3.1 Whole genome amplification labelling of clone DNA 55
2.6.3.2 Nick translation of clone DNA 57
2.6.3.3 DOP PCR labelling of flow sorted chromosomes 57
2.6.4 Pre-treatment of metaphase spreads and DNA fibre slides 58
2.6.5 Hybridisation of FISH slides 58
2.6.6 Detection of FISH slides 58
2.6.7 Microscope analysis of FISH slides 59
2.7 PCR 60
2.7.1 STS PCR 60
2.7.2 Colony PCR 61
2.7.3 LR PCR amplification of junction fragments 61
2.8 Sequencing 62
2.8.1 Sequencing of LR PCR products 62
2.8.2 End sequencing of fosmid clones 62
2.8.3 Full sequencing of fosmid clones 62
2.9 Custom-made library production 63
2.9.1 In-house cosmid library 63
2.9.1.1 Extraction of high molecular weight DNA 63
2.9.1.2 Preparation of vector arms 63
2.9.1.3 Partial digestion of DNA 64
2.9.1.4 Recovery of digested DNA 65
2.9.1.5 Ligation of vector arms to digested DNA 66
2.9.1.6 Packaging of ligated DNA 66
2.9.1.7 Preparation of plating cells 66
2.9.1.8 Plating of library 66
2.9.2 EpiFOS fosmid library 67
2.9.2.1 Size modification of DNA 67
2.9.2.2 End-Repair of sheared DNA 67
2.9.2.3 Size selection of end-repaired DNA 67
2.9.2.4 Ligation of size-selected DNA 68
2.9.2.5 Preparation of E.Coli host strain 68
2.9.2.6 Packaging of ligated DNA 68
2.9.2.7 Plating of library 69
2.10 Generation of fosmid library filters 69
2.10.1 Printing of polygrid filters from 384 well plates 69
2.10.2 Replication of clones onto round filters from agar plates 69
2.10.3 Preparation of clone DNA on filters for hybridisation 69
2.11 Radioactive screening of fosmid library filters 70
2.11.1 Generation of probe for library screening 70
2.11.1.1 Generation of probe by Inter-ALU PCR 70
2.11.1.2 Generation of probe by STS PCR 71
2.11.2 Hybridisation of radioactive DNA probe to fosmid library filters 71
2.11.3 Washing and detection of fosmid library filters 71

3 **Results: Investigation of chromosome rearrangements by array painting, comparative genomic hybridisation and PCR analysis** 72
3.1 Introduction 72
3.2 Summary of methods 74
3.3 Results; Refinement of chromosome translocation breakpoints using custom-made fosmid microarrays 76
3.3.1 Testing the custom-made fosmid microarray 77
3.3.1.1 Self versus self hybridisation 77
3.3.1.2 769p hybridisation 80
3.3.2 Array painting analysis of patient derivative chromosomes on custom-made fosmid microarrays 84
3.3.2.1 Patient t(2;7)(q37.3;p15.1) 85
3.3.2.2 Patient t(3;11)(q21;q12) 90
3.3.2.3 Patient t(7;13)(q31.3;q21.3) 92
3.4 Further refinement of chromosome translocation breakpoints by STS PCR, PCR microarrays and oligonucleotide microarrays 94
3.4.1 Refinement of breakpoints by STS PCR 94
3.4.1.1 Patient t(2;7)(q37.3;p15.1) 94
3.4.1.2 Patient t(3;11)(q21;q12) 97
3.4.1.3 Patient t(7;13)(q31.3;q21.3) 99
3.4.1.4 Summary of STS PCR data 101
3.4.2 Refinement of breakpoints by custom-made PCR product microarrays 101
3.4.3 Refinement of breakpoints by custom-made oligonucleotide microarrays 103
3.5 Chromosome translocation breakpoint amplification by LR PCR 104
3.6 Sequence of junction fragments 106
3.7 Most rapid method of translocation breakpoint mapping 107
3.8 Investigation of further rearrangement around translocation breakpoints using fosmid microarrays 110
3.9 Investigation of additional imbalance in patient t(2;7)(q37.3;p15.1) 111
3.9.1 Delineation of chromosome 3 duplication breakpoints in patient t(2;7)(q37.3;p15.1) by microarray analysis 112
3.9.2 Amplification and sequence of chromosome 3 duplication junction in patient t(2;7)(q37.3;p15.1) 114
3.10 Conclusions 116

4 **Results: Bioinformatic investigation of rearrangement breakpoints** 118
4.1 Introduction 118
4.2 Breakpoints and phenotypes 118
4.2.1 Patient phenotypes (as published in (Gribble et al. 2005)) 119
4.2.1.1 Patient t(2;7)(q37.3;p15.1) phenotype 119
4.2.1.2 Patient t(3;11)(q21;q12) phenotype 119
4.2.1.3 Patient t(7;13)(q31.3;q21.3) phenotype 119
4.2.2 Direct disruption of a gene by the translocation breakpoint 119

4.2.2.1 Direct gene disruption in patient t(2;7)(q37.3;p15.1) 120
4.2.2.2 Direct gene disruption in patient t(3;11)(q21;q12) 121
4.2.2.3 Direct gene disruption in patient t(7;13)(q31.3;q21.3) 122
4.2.3 Translocation breakpoints and position effect 124
4.2.3.1 Position effect in patient t(2;7)(q37.3;p15.1) 124
4.2.3.2 Position effect in patient t(3;11)(q21;q12) 126
4.2.3.3 Position effect in patient t(7;13)(q31.3;q21.3) 128
4.2.4 Duplication in patient t(2;7)(q37.3;p15.1) and phenotype 129
4.3 Breakpoints and genome architecture 131
4.3.1 Analysis of sequence at breakpoints 131
4.3.1.1 Patient t(2;7)(q37.3;p15.1) 133
4.3.1.2 Patient t(3;11)(q21;q12) 133
4.3.1.3 Patient t(7;13)(q31.3;q21.3) 134
4.3.1.4 Summary of basepair sequence analysis at the breakpoints 134
4.3.2 Analysis of repeat structures and recombination motifs at breakpoints 134
4.3.2.1 Analysis of repeat structures around t(2;7)(q37.3;p15.1) breakpoints 136
4.3.2.2 Analysis of repeat structures around t(3;11)(q21;q12) breakpoints 137
4.3.2.3 Analysis of repeat structures around t(7;13)(q31.3;q21.3) breakpoints 137
4.3.2.4 Summary of sequence motifs found at translocation breakpoints 138
4.4 Mechanisms underlying genomic rearrangements 138
4.4.1 Translocations in patients t(2;7)(q37.3;p15.1), t(3;11)(q21;q12) and t(7;13)(q31.3;q21.3) 138
4.4.2 Duplication in patient t(2;7)(q37.3;p15.1) 139
4.5 Conclusions 140

5 **Results: Investigation of chromosome translocations using custom-made libraries** 141
5.1 Introduction 141
5.2 Comparison of custom-made library protocols 141
5.2.1 Cosmid library 142
5.2.2 EpiFOS fosmid library 143
5.2.2.1 Using whole genomic DNA 143
5.2.2.2 Using flow sorted derivative chromosomes 145
5.2.3 Conclusions 147
5.3 Using the EpiFOS kit to generate a library from a patient with a characterised translocation as a test case for developing screening methods 148
5.4 Comparison of screening methods using gridded library fosmid clones 148
5.4.1 Inter ALU PCR screening 149

5.4.2 PCR product pools 150
5.4.3 End sequencing 150
5.4.4 Specific breakpoint PCR product 153
5.4.5 Summary of fosmid clones identified by the different screening methods 154
5.4.6 Conclusions 157
5.4.7 Screening of round library plates 160
5.5 . Mapping translocation breakpoints using a custom-made fosmid library from a patient with a t(2;6)(q21.1;q25.1) translocation 162
5.5.1 Isolation of derivative chromosomes by flow sorting 163
5.5.2 Verification of flow sorted derivative chromosomes by reverse chromosome painting. 164
5.5.3 Mapping translocation breakpoints by array painting 165
5.5.4 Isolation of breakpoint spanning fosmid clones using a custom-made library 167
5.5.4.1 Production of the t(2;6)(q21.1;q25.1) fosmid library 167
5.5.4.2 Screening of the t(2;6)(q21.1;q25.1) library using PCR product pools. 168
5.5.5 Identification of translocation breakpoints by sequence analysis 169
5.6 Conclusions 170

6 **Results: Investigation of viral integration using microarrays and custom-made libraries** 172
6.1 Introduction 172
6.1.1 The HHV-6 virus structure 172
6.1.2 Patient details 173
6.1.2.1 Flow karyograms 174
6.2 FISH analysis of viral integration sites in 3 patients 175
6.3 Investigation of integration using genomic microarrays 177
6.3.1 Verification of the microarray 177
6.3.2 Comparison between the 3 patients 178
6.4 Investigation of integration using a custom-made library 179
6.4.1 Generation of custom-made library 179
6.4.2 Screening of custom-made library 180
6.5 Investigation of integration using FISH onto extended DNA fibres 183
6.6 Conclusions 184

7 **Discussion** 185
7.1 Methods of rearrangement breakpoint mapping 185
7.1.1 Breakpoint mapping using microarray based techniques 185
7.1.2 Translocation breakpoint mapping using custom-made fosmid libraries 188
7.1.3 Application of techniques developed for the mapping of rearrangement breakpoints to the investigation of viral integration in the human genome 190

7.2 Bioinformatic analysis of genomic rearrangement breakpoints 191
7.2.1 Genotype to phenotype correlations 192
7.2.1.1 Identification of non-pathological variation 192
7.2.1.2 Gene disruption 193
7.2.2 Mechanisms underlying genomic rearrangements 195
7.2.2.1 Predisposition 196
7.2.2.2 Occurrence of rearrangements 196
7.3 Recent developments and the future of breakpoint mapping 197
7.4 Conclusions 199

1 Introduction

1.1 The discovery of human chromosomes

Human chromosomes were first observed by Flemming and Arnold in the 1880's with the correct number of chromosomes in a human cell finally identified as 46 in 1956 (Ford and Hamerton 1956; Tijo and Levan 1956). Human chromosomes are morphologically distinct, and were easily classified into 7 groups (a to g) according to their size and relative positions of the centromeric constriction (Patau 1960). In 1968 Torbjorn Caspersson developed a method of staining human metaphase chromosomes reproducibly to give a distinctive pattern of dark and light bands along their lengths (Caspersson et al. 1968; Caspersson et al. 1972). This distinctive banding pattern along with the relative sizing and centromeric positions are still used in general karyotyping analysis and the identification of chromosome rearrangements today.

1.2 Chromosome rearrangements

The development of chromosome banding allowed much more detailed analyses of chromosomes such that structural rearrangements were rapidly identified. Structural chromosome abnormalities including translocations, deletions, amplifications and inversions arise from recombination between chromosomes or the misrepair of chromosome breaks. These chromosome rearrangements are often identified by cytogenetic techniques, initially by studying the patient's karyotype using chromosome size and banding patterns, and more recently using microarrays. Chromosome rearrangements have been associated with abnormal phenotypes; either by directly disrupting gene structures, modifying gene regulation or alteration of the copy number.

1.2.1 Chromosome translocations

One of the first translocations identified was the pathogenic somatic translocation implicated in Chronic Myeloid Leukaemia (Rowley 1973). The t(9;22)(q34;q11)

reciprocal translocation produces the characteristic Philadelphia marker chromosome, which disrupts chromosome 9 within the Abelson murine leukaemia viral oncogene homologue 1 (*ABL*) gene and chromosome 22 within the breakpoint cluster region (*BCR*) gene. The resulting gene fusion product, BCR-ABL, inhibits DNA repair and accelerates cell division.

Constitutional translocations arise very early in development; either as the result of an abnormal gamete or possibly after abnormal fertilisation or early embryo formation and affect every cell within a body. Two classes of constitutional translocations exist: balanced (with no accompanying net gain or loss of material) and unbalanced. Robertsonian translocations are a specific class of unbalanced translocation resulting in the formation of a dicentric derivative chromosome after the joining of 2 acrocentric chromosomes (parts of the acrocentric arms are lost). Both balanced reciprocal translocations and Robertsonian translocations are stable through mitosis (Figure 1.1).

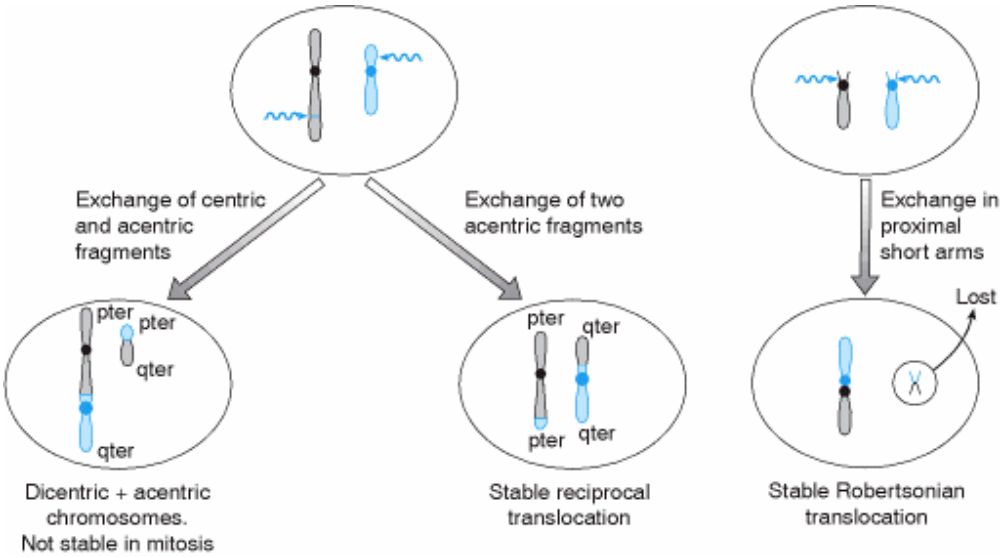


Figure 1.1 The origins of translocations (Strachan and Read 1999). Robertsonian translocations are a result of exchanges between the proximal short arms of the acrocentric chromosomes 13, 14, 15, 21 and 22. Both centromeres are present but function as one, so the derivative chromosome is stable.

The most common recurrent non-Robertsonian constitutional translocation is t(11;22)(q23;q11). It is associated with an increased frequency of spontaneous abortion and a 10 fold increase in risk of breast cancer within carriers. The 11q23 breakpoints fall within an AT-rich repeat (Edelmann et al. 1999) and the 22q11 breakpoints fall within a Low Copy Repeat (LCR22) (Funke et al. 1999) suggesting that the mechanism by which the translocation occurs is the occurrence of double strand breaks within palindromic sequences leading to illegitimate recombination events (Edelmann et al. 2001).

1.2.1.1 Frequency of translocations

It is estimated that 1 in 200 live births have an apparently balanced translocation and 1 in 500 have an unbalanced translocation (Jacobs et al. 1992). The frequency of *de novo* reciprocal apparently balanced translocations is approximately 1 in 2000 (Warburton 1991). Many carriers of a balanced translocation appear phenotypically normal and balanced translocations often go unnoticed until segregation of an unbalanced form of the translocation results in recurrent miscarriages or offspring with congenital abnormalities. Only 25% of gametes produced from a cell with a reciprocal translocation will contain copies of normal chromosomes, a further 25% will carry both derivative chromosomes and the other 50% will carry a partial monosomy for one chromosome and a partial trisomy for the other chromosome which may be lethal (Figure 1.2).

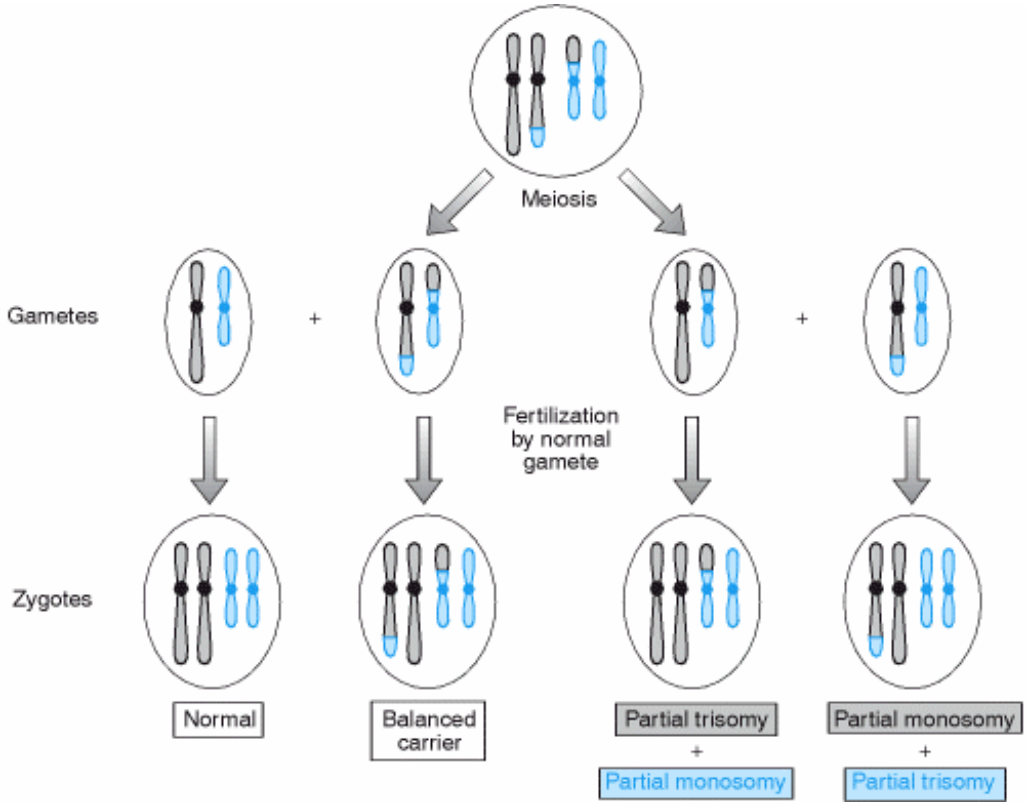


Figure 1.2 Results of meiosis in a carrier of a balanced translocation (Strachan and Read 1999).

The mapping and sequencing of chromosome breakpoints will improve our understanding of the mechanisms underlying chromosomal rearrangements and their involvement in disease.

To date, the sequence across the junctions of less than 50 non-recurrent constitutional translocations have been published (Table 1.1). Increasing the pool of data by developing efficient methods of breakpoint mapping may help to elucidate the mechanism underlying the formation of chromosomal rearrangements.

Translocation	Reference
t(X;21)(p21;p12)	Bodrug et al., 1987
t(X;2)(p21;q37)	Bodrug et al., 1991
t(X;4)(p21;q35)	
t(X;1)(p21;p34)	Cockburn, 1991
t(X;4)(p21.2;q31.22)	Giacalone and Francke, 1992
t(4;22)(q12;q12.2)	Arai, Ikeuchi and Nakamura, 1994
t(2;22)(q14;q11.21)	Budarf et al., 1995
t(X;5)(p21;q31.1)	van Bakel et al., 1995
t(X;9)(p21.1;q34.3)	Toriello et al., 1996
t(21;22)(p12;q11)	Holmes et al., 1997
t(X;8)(p22.13;q22.1)	Ishikawa_Brush et al., 1997
t(17;22)(q11.2;q11.2)	Kehrer-Sawatzki et al., 1997
t(6;7)(q16.2;p15.3)	Krebs et al., 1997
t(8;17)(p11.2;p13.3)	Kurahashi et al., 1998
t(1;10)(p22;q21)	Roberts, Chernova and Cowell, 1998
t(2;19)(q11.2;q13.3)	Yoshiura et al., 1998
t(6;12)(q16.2;q21.2)	Ikegawa et al., 1999
t(1;6)(p22.1;q16.2)	Holder, Butte and Zinn, 2000
t(1;8)(q21.1;q22.1)	Matsumoto et al., 2000
t(1;11)q42.1;q14.3)	Millar et al., 2000
t(12;22)(q24.1;q13.3)	Bonaglia et al., 2001
t(1;19)(q21.3;q13.2)	Nothwang et al., 2001
t(9;11)(p24;q23)	Willett-Brozick et al., 2001
t(7;16)(q11.23;q13)	Duba et al., 2002
t(1;8)(p34.3;q21.12)	McMullan et al., 2002
t(2;8)(q31;p21)	Spitz et al., 2002
t(2;8)(q31;p21)	Sugawara et al., 2002
t(6;13)(q21;q12)	Vervoort et al., 2002
t(7;22)(p13;q11.2)	Hill et al., 2003
t(6;11)(q14.2;q25)	Jeffries et al., 2003
t(4;22)(q35.1;q11.2)	Nimmakayalu et al., 2003
t(X;7)(p11.3;q11.21)	Shoichet et al., 2003
t(1;7)(q41;p21)	David et al., 2003
t(1;22)(p21.2;q11)	Gotter et al., 2004
t(3;8)(p14.2;q24.2)	Rodriguez-Perales et al., 2004
t(2;6)(q24.3;q22.31)	Bocciardi et al., 2005
t(4;17)(q28.3;q24.3)	Velagaleti et al., 2005
t(1;7)(p22;q32)	Borg et al., 2005
t(4;15)(q27;q11.2)	Schule et al., 2005
t(4;15)(q22.3;q21.3)	Klar et al., 2005
t(9;11)(q33.1;p15.3)	Tagariello et al., 2006
t(6;17)(p21.31;q11.2)	Mansouri et al., 2006
t(5;14)(q21;q32)	Haider et al., 2006
t(17;22)(q21.1;q12.1)	Gribble et al., 2007
t(2;7)(q37.1;q36.3)	
t(11;17)(p13;p13.1)	
t(2;7)(q37.1;q21.3)	Bocciardi et al., 2007

Table 1.1 Summary of published non-recurrent constitutional translocations.

1.2.1.2 Associated phenotypic risk

A study into the outcome of 377,357 amniocenteses at birth showed that the risk of a *de novo* apparently balanced translocation having an associated phenotypic abnormality is 6.1% (Warburton 1991). The risk of congenital abnormalities amongst carriers of balanced translocations is double that seen amongst individuals with normal karyotypes (Warburton 1991). Analysis of rearrangement breakpoints in relation to the banding pattern observed along Giemsa banded metaphase chromosomes showed that 84% of breakpoints occurred in Giemsa negative, gene rich regions of the genome (Warburton 1991; Niimura and Gojobori 2002).

A recent study investigated the hypothesis that translocation breakpoints in normal individuals were simple and did not disrupt genes (Baptista et al. 2005). The breakpoints in 13 phenotypically normal individuals with apparently balanced translocations were mapped by FISH. At the resolution of the study (approximately 150Kb) the breakpoints were seen to directly disrupt a gene in 2 patients and possibly disrupt a gene in a further 8 patients, the significance of which remained undetermined. An additional observation was made that the translocation breakpoints in phenotypically normal patients were not accompanied by additional imbalance in contrast to the breakpoints in phenotypically abnormal patients which are often seen to be accompanied by cryptic imbalances. FISH and PCR studies have reported that 8 out of 30 apparently balanced reciprocal translocations are more highly rearranged than identified by G-banding (Kumar et al. 1998; Astbury et al. 2004; Patsalis et al. 2004). We have similarly shown that 6 out of 10 apparently balanced translocations as described by G-banding were more complex when analysed by DNA microarray analysis (Gribble et al. 2005). A different study using microarrays and FISH has shown that of the 4 translocation patients studied, all were more complex than originally thought and that in 3 of the patients the

abnormal phenotype was postulated to be associated with the additional imbalance (Ciccone et al. 2005). These studies have led to the suggestion that the genetic diagnosis of constitutional rearrangements should routinely include molecular karyotyping by array based whole genome screening in order to detect submicroscopic imbalances (Vermeesch et al. 2007).

Many published translocation breakpoints have shown that the direct disruption of a gene can lead to an associated phenotype (Bhalla et al. 2004; Bocciardi et al. 2005; Klar et al. 2005) but it is also thought that translocation breakpoints can cause an affect on genes several Kb away. This position effect phenomenon was initially described in *Drosophila* and yeast and was first documented in humans in 1995. A study into a patient with a t(4;11)(q22;p13) translocation showed that the translocation breakpoints did not directly disrupt the *PAX6* gene to which the aniridia phenotype was associated in patients with deletions of the same region (Fantes et al. 1995). The translocation breakpoint was mapped approximately 125-185Kb distal to the *PAX6* gene in the patient with aniridia, suggesting that the translocation was still exhibiting an effect on the gene and the patient's phenotype. Other examples of position effect arising from a chromosome rearrangement include a t(2;8)(q31;p21) translocation which affects the *HOXD* gene 60Kb away from the chromosome 2 breakpoint in a patient with Mesomelic Dysplasia and vertebral defects (Spitz et al. 2002) and a t(6;11)(q14.2;q25) translocation affecting the *B3GAT1* gene which lies 299Kb centromeric to the chromosome 11 breakpoint in a patient with psychosis (Jeffries et al. 2003). The longest range position effect observed to date was observed in a patient with a t(4;7;8;17) translocation. The patient presented with Campomelic Dysplasia which has been attributed to the *SOX9* gene on chromosome 17q24.3. The chromosome 17 breakpoint was found to fall 1.3Mb downstream of the gene (Velagaleti et al. 2005).

The position effect phenomenon arises from the disruption of *cis*-acting regulatory elements such as promoters, enhancers and silencers which can be directly altered, distanced from the gene they influence, or brought into proximity of a gene not normally under their control when a chromosome undergoes a rearrangement. These elements have been observed as far away as 1.1Mb from the gene they regulate, as in the case of the SOX9cre1 element which was identified upstream of the *SOX9* gene (Bien-Willner et al. 2007).

1.2.2 Deletions, duplications, inversions

Regions of repeat sequence have been implicated in the mechanism of rearrangement such as deletions, duplications and inversions as summarised in Figure 1.3. Homologous regions within the genome are thought to align, allowing recombination to occur between the sequences, resulting in rearrangements.

A study of 14.677 conceptions has estimated the frequency of unbalanced structural rearrangements at 1 in 460 (Jacobs et al. 1992) and a study of 377,357 amniocentesis estimated the rate of inversions to be 1 in 10,000 with a 9.4% risk of an associated congenital abnormality (Warburton 1991).

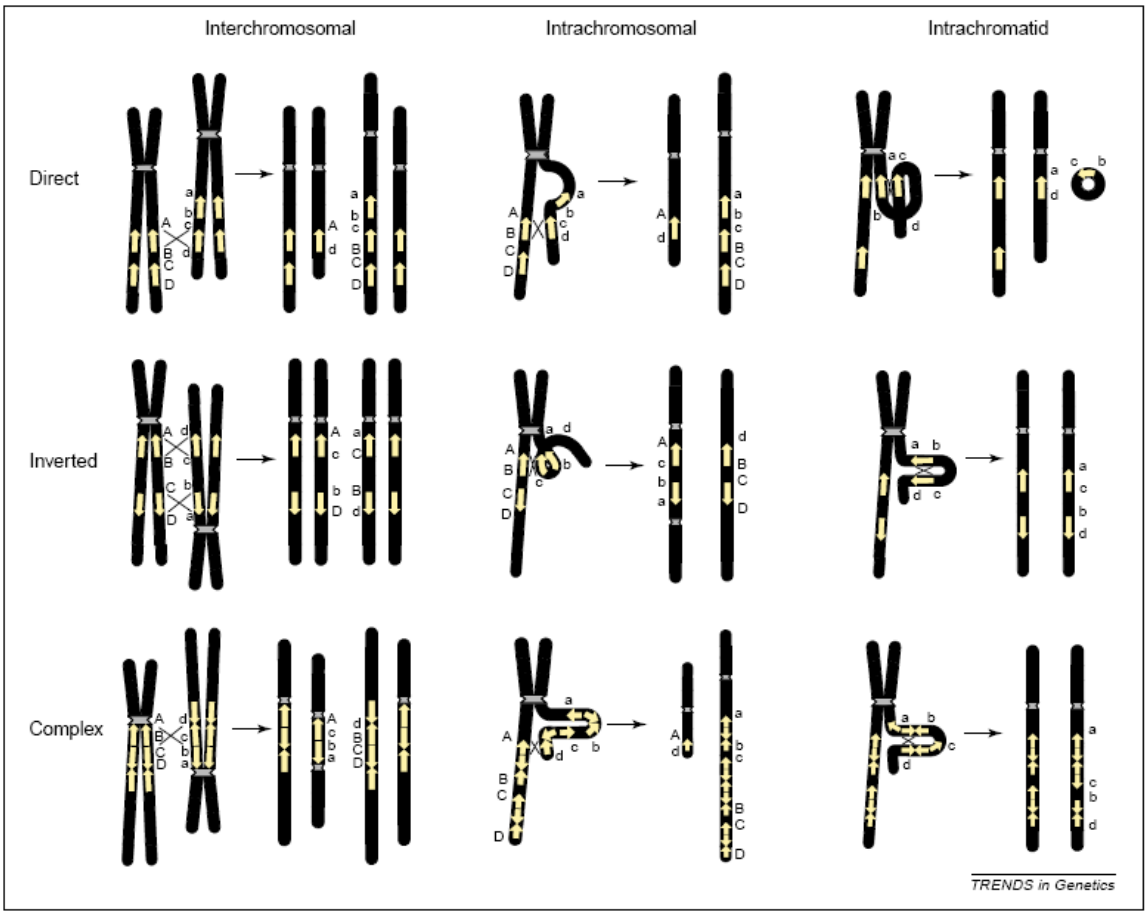


Figure 1.3 Schematic representation of homologous recombination based mechanisms for genomic rearrangements (Stankiewicz and Lupski 2002). Yellow arrows depict the orientation of regions of homology.

Estimates of the repeat content of the genome range from approximately 5-50%. One physical mapping study using 1,243 randomly selected BAC clones found that 5.4% hybridised to more than one chromosomal location (Cheung et al. 2001). Bioinformatic analysis of the human genome draft sequence estimated the repeat content to be approximately 50% (Lander et al. 2001). Repeatmasker analysis of NCBI Build 36 of the human genome reference sequence reveals that current calculations estimate that 48.9% of the genome is repeat sequence (data extracted from http://genome.ucsc.edu/). The discrepancy between the estimates

of repeat sequence can be explained by the resolution afforded by the analysis techniques. Physical analysis of the genome by FISH has a limited resolution and relies on a region of hybridisation being detected visually. Bioinformatic tools such as Repeatmasker will consider repeats of any size ranging from several Megabases (e.g. segmental duplications) to only a few basepairs (e.g. simple repeats) which would not be detected visually by FISH.

1.2.2.1 Phenotypic effect of deletions, duplications and inversions

Changes in the number of copies of an expressed gene may alter a patient's phenotype unless dosage compensation is observed. Loss of a copy of a gene may result in haploinsufficiency, or an increase in the number of genes may result in overexpression. All these changes refer to loss or gain of a functional gene with its associated regulatory elements. However, the breakpoints associated with deletions, duplications and inversions may also disrupt a gene as discussed in section 1.2.1.2.

1.2.2.2 Genomic disorders

The term "genomic disorders" was coined to describe a group of recognised diseases that arise due to rearrangements in the genome which are mediated by the genomic architecture. Well studied examples include Sotos syndrome, Charcot-Marie-Tooth disease type 1A (CMT1A), hereditary neuropathy with liability to pressure palsies (HNPP), Smith Magenis Syndrome (SMS) and dup(17)(p11.2p11.2) syndrome.

Sotos syndrome is a genomic disorder characterised by childhood overgrowth, mental retardation and specific craniofacial features associated with haploinsufficiency of the *NSD1* gene on 5q35 either arising from deletion of the gene or a mutation. Analysis of the Sotos syndrome region revealed two complex mosaic low copy repeats, either side of the *NSD1* gene. The proximal repeat of approximately 390Kb and the distal repeat of approximately 429Kb show an

overall homology of 98.5%. Studies have shown that the deletions associated with Sotos syndrome are likely to be the result of rearrangements between these low copy repeats (Kurotaki et al. 2005; Visser et al. 2005).

CMT1A and HNPP are dysmyelinating peripheral neuropathies resulting from altered dosage of the *PMP22* gene caused by reciprocal duplication and deletion events on 17p12. These recurrent rearrangements of a 1.4Mb genomic region are mediated by proximal and distal low copy repeats which are 24Kb in length with 99% homology (Reiter et al. 1997; Shaw et al. 2004)

SMS is a syndrome associated with mental retardation and multiple congenital abnormalities whilst dup(17)(p11.2p11.2) syndrome sometimes results in milder forms of mental retardation. The 4Mb and 5Mb deletions associated with SMS and the reciprocal duplications associated with dup(17)(p11.2p11.2) syndrome are mediated by 3 LCRs (256Kb proximal, 241Kb middle, 176Kb distal) with 98% homology (Smith et al. 1986; Potocki et al. 2000; Park et al. 2002).

In all these cases, it is believed that the rearrangements arise because of illegitimate homologous recombination via the repeat structures identified. Because of this mechanism, the disorders are seen to be recurrent within the human population, although the exact breakpoints of the rearrangements are observed to vary slightly. For example, a study of patients with CMT1A found that 19 out of 24 breakpoints fell within a 741bp region within the CMT1A repeat (Lopes et al. 1998).

1.2.3 Analysis of chromosome rearrangements

Many techniques have been applied to the analysis of chromosome rearrangements. Conventional cytogenetic techniques such as G-banding of metaphase chromosomes are generally applied for the initial identification of rearrangements, but more recently microarray based techniques have been

developed. Analysis of a chromosome's size and constitution by flow cytometry can also be used to identify large aberrations. Rearrangement breakpoints can be investigated at increased resolution using FISH, somatic cell hybrids, array painting and custom-made libraries.

1.2.3.1 Metaphase chromosome banding analysis

Many stains can be used to analyse banding patterns along metaphase chromosomes. For example, the fluorescent stain DAPI preferentially binds to AT rich DNA resulting in regions of dark and light staining along the chromosome corresponding to GC and AT rich regions (Lin et al. 1977). Conventional G-banding uses Giemsa to produce a banding pattern. Analysis of this pattern in relation to the human genome reference sequence showed that Giemsa positive bands (dark) were AT rich and gene poor and Giemsa negative bands (light) were GC rich and gene rich (Niimura and Gojobori 2002). Analysis of chromosome banding patterns can identify chromosome rearrangements and localise the breakpoints to within approximately 3Mb depending on the quality and length of the prepared chromosomes (Lichter et al. 2000).

1.2.3.2 FISH

The development of fluorescent *in situ* hybridisation (FISH) has increased the resolution with which chromosome rearrangements can be studied and the completion of the Human Genome Project has provided an invaluable resource of fully sequenced large insert clones enabling scientists to study human chromosomes in more detail more easily. Once a chromosome rearrangement has been identified by banding analysis, large insert clones can be selected across the region of interest, fluorescently labelled and hybridised to metaphase chromosomes (Bauman et al. 1980). Analysis of the signals obtained from the hybridisation of subsequent clones along the chromosome can identify clones which span the rearrangement breakpoints.

1.2.3.3 Fluorescence activated chromosome sorting

Fluorescence activated chromosome sorting (FACS) was developed in 1979 as a technique for the separation of individual human chromosomes (Gray et al. 1979). Two fluorochromes were used for chromosome staining; Hoechst 33258 which binds to AT rich DNA and Chromomycin A3 which binds to GC rich DNA. After staining, chromosomes can be resolved by their DNA content and relative size with the exception of chromosomes 9-12 which remain as a single peak due to their similar size and basepair constitutions (Figure 1.4). Rearrangements such as reciprocal translocations which alter the size and basepair constitution of chromosomes can be detected using flow cytometry.

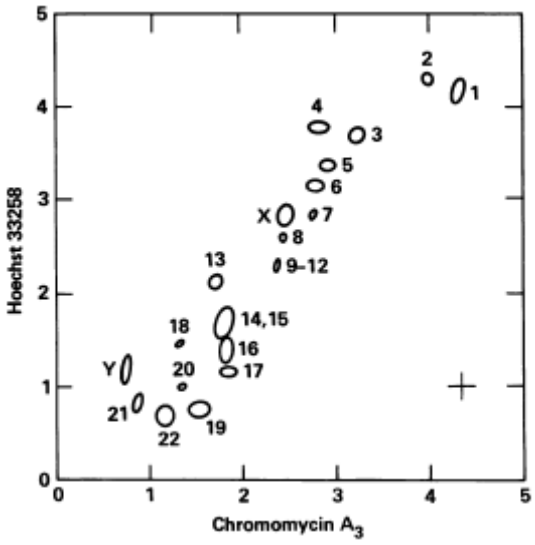


Figure 1.4 Statistical summary of peaks seen for the resolution of human chromosomes when stained with Hoechst 33258 and Chromomycin A3 for a normal 46,XY karyotype (Langlois et al. 1982). Chromosomes 9, 10, 11 and 12 do not resolve into discrete peaks.

Chromosome sorting technologies can be used for the isolation of derivative chromosomes and to improve the purity of template DNA for the analysis of chromosome rearrangements.

1.2.3.4 Somatic cell hybrids

Somatic cell hybrids have been used to isolate derivative chromosomes from patients with chromosome translocations. The derivative chromosomes can be digested to create a restriction digest map for Southern blotting.

1.2.3.5 Southern blotting

Southern blotting has been used to identify fragments of DNA that contain translocation breakpoints. Fragments of DNA from the digested derivative chromosomes isolated by somatic cell hybrids, or digested genomic DNA from the patient can be screened using probes designed from breakpoint spanning clones identified by FISH. These fragments can then be cloned, sequenced and analysed in comparison with the human genome reference sequence to identify the breakpoint position (Vervoort et al. 2002).

1.2.3.6 Comparative genomic hybridisation

The principle of comparative genomic hybridisation (CGH) was developed to analyse the genomic changes within cancer cell lines in direct comparison with the DNA of normal individuals (Kallioniemi et al. 1992). DNA from a test cell line labelled with one fluorochrome, and a reference DNA labelled with a second fluorochrome were simultaneously hybridised to metaphase chromosomes in the presence of Cot1 DNA to block repetitive sequences. Analysis of the changes in the ratios of intensity of the 2 fluorochromes revealed regions of amplification and deletion within the 2 samples at a resolution of approximately 10Mb with subsequent studies increasing the resolution to approximately 3Mb (Kirchhoff et al. 1999).

In order to increase the resolution of amplification and deletion screening within the genome of interest further, genomic microarrays were developed and used as a target for CGH, replacing metaphase chromosomes (Solinas-Toldo et al. 1997).

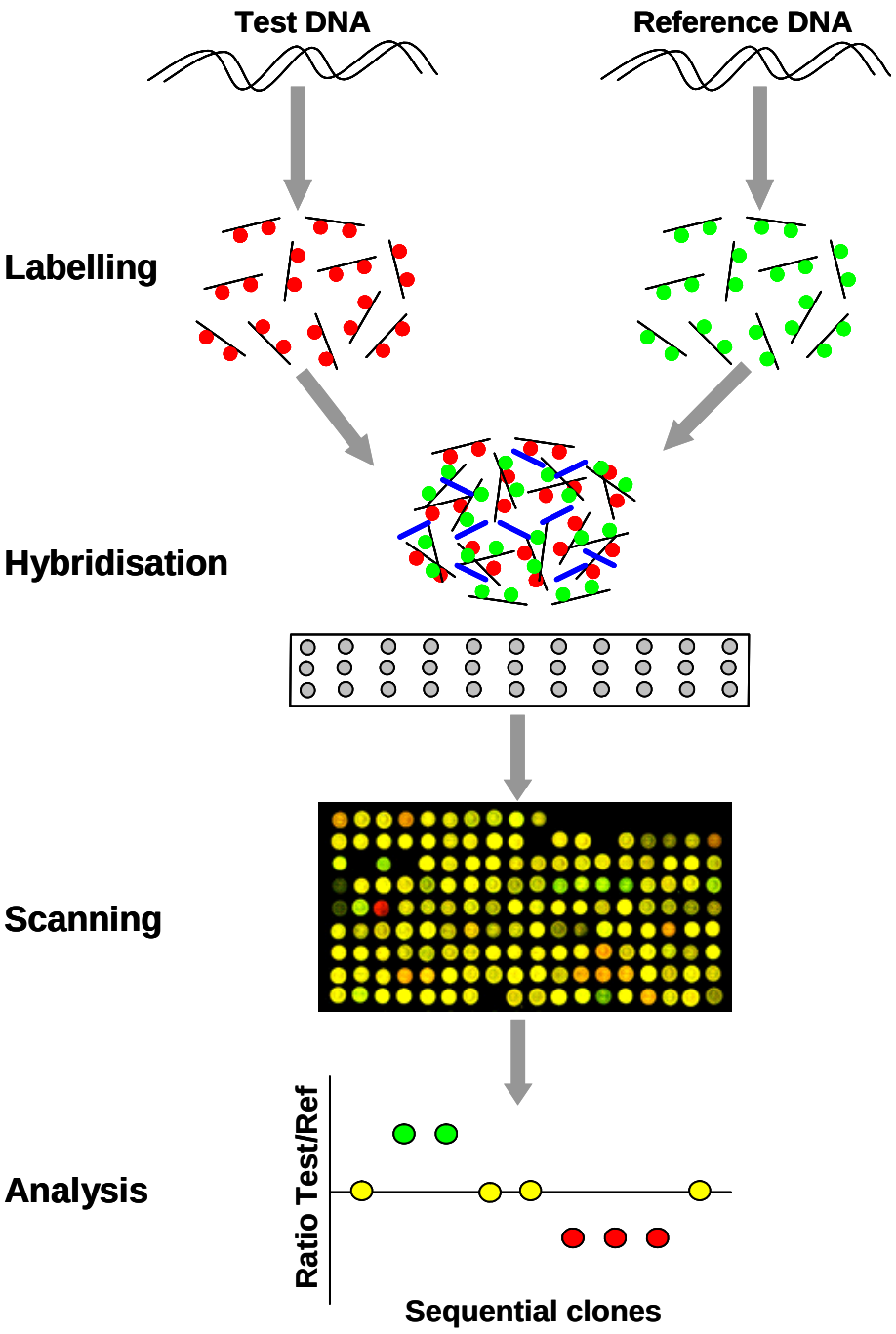


Figure 1.5 The principles of array CGH. The test (red) and reference (green) DNAs are differentially labelled and hybridised to a genomic microarray slide in the presence of Cot1 DNA (blue). Amplifications and deletions are identified by differences in the fluorochrome ratios.

During an array CGH experiment, patient genomic DNA and reference DNA are differentially labelled and competed with Cot1 DNA. The mixture is then hybridised to the target DNA spotted on the microarray slide to identify any genomic imbalances between the 2 samples. A change in the ratio of fluorescence intensities of the clones indicates a copy number change (Figure 1.5).

The advent of genomic microarrays has meant that a patient's genome can be studied more rapidly and at a greater resolution than was previously possible by FISH analysis with the resolution of the genomic microarray limited only by the size and spacing of the probes (clones, PCR products or oligonucleotides) selected during its design.

The most frequently used DNA microarrays to date have a resolution of approximately 1-1.4Mb (Pollack et al. 1999; Snijders et al. 2001; Fiegler et al. 2003a; Vissers et al. 2003; Schoumans et al. 2004; Shaw-Smith et al. 2004). These arrays provide genome-wide scans at relatively low resolution. To further refine rearrangement breakpoints, higher resolution arrays can be used. Arrays at tiling path resolution have recently been created with the ability to map chromosome breakpoints within a spanning clone in a single experiment (Ishkanian et al. 2004; Fiegler et al. 2007). The large insert clones used to create these arrays provide a resolution of approximately 150Kb.

To generate genomic clone microarrays with further increased resolution, alternative library resources are available. A library of fosmid clones was created and end sequenced as part of the human genome project providing a validation method for the sequence data generated (IHGSC 2004). This library had on average an 8 fold coverage of the euchromatic regions of the human genome providing a high level of redundancy. Clones can be selected from this publicly available resource to cover a genomic region of interest and create a targeted

custom-made microarray with a high level of redundancy and at a higher resolution than that afforded by large insert clone microarrays.

Custom-made microarrays using alternative targets to genomic clones have been created to investigate particular genomic regions of interest at increased resolution. Microarrays constructed from PCR products were used to analyse genomic imbalances across the Neurofibromatosis 2 gene. PCR products ranging in size from 150-650bp were pooled affording a resolution of 23Kb (Mantripragada et al. 2003). Use of PCR products spotted as individual targets onto microarrays slides has subsequently increased the resolution of PCR product arrays (Dhami et al. 2005). By amplifying genomic sequence using PCR primer pairs tagged with an 8bp universal sequence and performing a second round of amplification using this universal sequence to attach an amino group, each product was able to be spotted individually onto the microarray, allowing the resolution of the microarray to be as small as the PCR product.

A single nucleotide polymorphism (SNP) microarray was first developed to analyse sequence polymorphisms within genomes (Chee et al. 1996). 135,000 oligonucleotide probes covering the 16.6Kb human mitochondrial genome were synthesised and spotted onto an array, revolutionising the resolution with which genomes could be studied.

Oligonucleotide based microarrays have since been developed to investigate imbalance within genomes, so increasing the resolution of array CGH. A study using a microarray with 60-mer oligonucleotides showed that when compared with CGH using BAC arrays the oligonucleotide arrays were able to detect amplifications with a higher accuracy and greater special resolution (Carvalho et al. 2004).

Microarrays using oligonucleotides directly synthesised onto microarray slides have been used to probe the genomes of 5 cancer patients at a resolution of 6Kb (Selzer et al. 2005). In the same study, custom-made oligonucleotide microarrays at higher resolution were created to map the breakpoints of the rearrangements identified by the whole genome screen to intervals as low as 50bp. NimbleGen Systems, Inc. routinely use oligonucleotide microarrays with a median probe spacing of 713bp for whole genome screening. Custom-made oligonucleotide microarrays can also be generated with a median probe spacing from 10bp.

The power of array CGH for the identification of rearrangements involved in disorders has been demonstrated by a study of 290 individuals with mental retardation (Sharp et al. 2006). A genomic BAC microarray was created to cover regions of the genome suspected of instability due to flanking segmental duplications. Four individuals were identified with the same 500Kb deletion of 17q21.31 and subsequent analysis of these patients using a targeted oligonucleotide microarray with a mean spacing of 1 probe every 131bp revealed a minimal critical region of 478Kb.

1.2.3.7 Array painting

Array CGH is a powerful tool for identifying unbalanced chromosome rearrangements, but it is unable to detect balanced rearrangements. Array painting is an adaptation of the array CGH technique which overcomes this limitation. The genomic DNA used in CGH experiments is replaced with flow sorted derivative chromosomes which are differentially labelled and hybridised to the microarray to map translocation chromosome breakpoints (Fiegler et al. 2003b) (Figure 1.6).

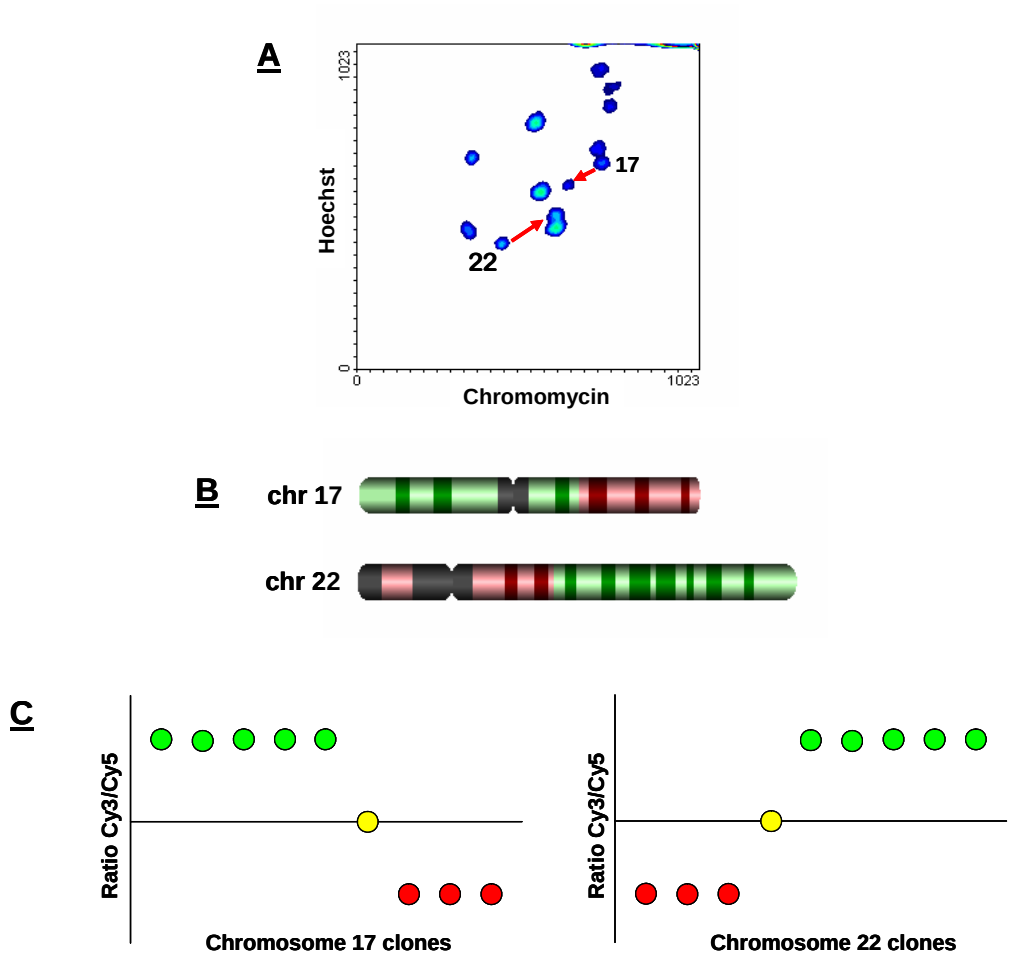


Figure 1.6 ***Schematic showing the process of array painting for a patient with a t(17;22)(q21.1;q12.2) translocation. A Partial flow karyogram showing the shift between chromosome 17 and derivative chromosome 17 and also chromosome 22 and derivative chromosome 22 (indicated by the red arrows); B Ideogram showing the translocation after differential labelling of the flow sorted derivative chromosomes; C Array painting analysis of the Cy3/Cy5 ratios. The breakpoint region is identified by a shift from a low to high ratio or vice versa. A spanning clone (yellow) will show an intermediate ratio.***

Array painting in conjunction with array CGH can map the breakpoints of reciprocal translocations and detect any additional imbalance within the patient's genome.

1.2.3.8 Custom-made libraries

As discussed in section 1.2.1.2, apparently balanced reciprocal translocations are often more complex than initially identified by G-banding, highlighting the need for alternative methodologies of breakpoint mapping that do not rely heavily on knowledge of the sequence surrounding the breakpoints.

The use of libraries has been instrumental to the mapping and sequencing of the human genome. Large insert clone libraries were constructed; YAC clones (Traver et al. 1989; Albertsen et al. 1990), PAC clones (Ioannou et al. 1994), BAC clones (Kim et al. 1996; Osoegawa et al. 2001), cosmid clones (Wood et al. 1992) and fosmid clones (IHGSC 2004). Generation and screening of a custom-made library using DNA from a translocation patient relies only on limited sequence information surrounding the breakpoints.

During the generation of a library, DNA is fragmented to the appropriate size, ligated into a vector, packaged into a suitable host and cloned as summarised in Figure 1.7. Many different vector and host systems exist, such as yeast artificial chromosomes (YACs), P1 derived artificial chromosomes (PACs), bacterial artificial chromosomes (BACs), fosmids, cosmids and plasmids capable of containing inserts ranging in size from approximately 1Mb to under 5Kb.

Many factors affect the decision of which vector and host system is used. YAC libraries were initially used in cloning studies as they were capable of cloning large fragments of DNA up to 1Mb, however they have been shown to be prone to rearrangement. The choice of cloning vector used is often dependent upon the resolution of the library required; fosmid or cosmid clones contain inserts of approximately 40Kb and plasmid clones contain inserts less than 15Kb in size.

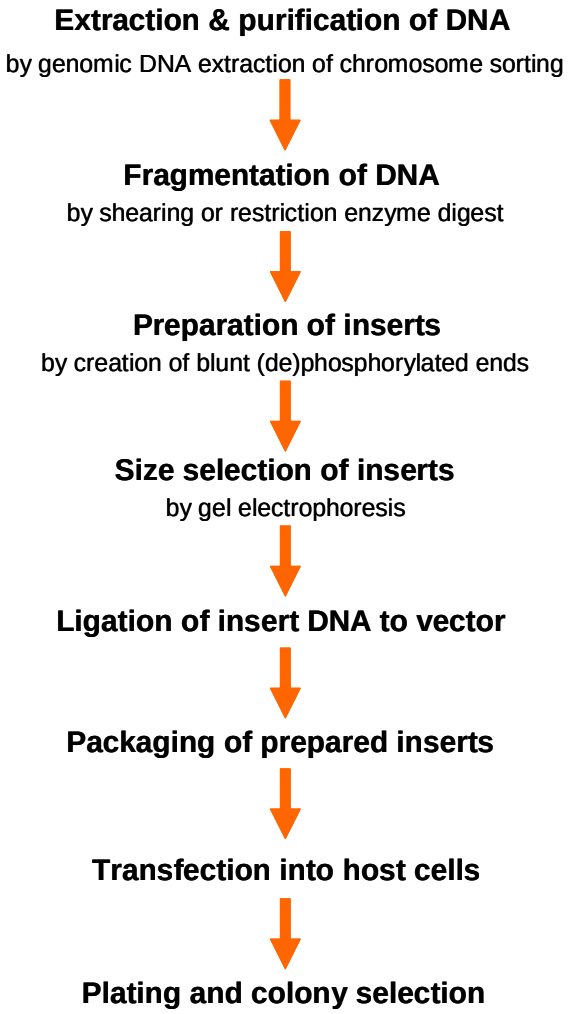


Figure 1.7 Schematic showing the key stages in the production of a custom-made library. Precise stages required are dependant upon the type of vector and host.

1.2.3.9 Sequence analysis

The completion of the human genome reference sequence announced in 2004 provided an invaluable resource for the investigation of chromosome rearrangements at sequence level (IHGSC 2004). Approximately 99% of the euchromatic regions of the genome had been sequenced, with the data made publicly available in genome browsers such as UCSC (http://genome.ucsc.edu/)

and Ensembl (http://www.ensembl.org/Homo_sapiens/index.html). In addition to the reference sequence, these databases provided details of genomic clones and their relative positions available for genomic research. These websites also allowed researchers to download sequence and interrogate regions of interest for fully annotated genomic structures including genes, repeat elements, polymorphic regions and regions of conservation. Multiple web-based tools also exist for the analysis of sequence for genomic motifs (Appendix A1). The analysis of genomic sequence with relation to rearrangement breakpoints and their influence in disease is enhanced by the availability of such tools which enable scientists to link genomic aberrations to phenotypes, therefore identifying candidate disease genes.

1.2.4 Public databases of constitutional rearrangements

In general, constitutional non-recurrent rearrangements are investigated by clinical and research groups around the world, with little or no contact. In an effort to enhance communication between these groups, web-based databases have been developed such as the Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources (Decipher; http://www.sanger.ac.uk/PostGenomics/decipher/), the Gross Rearrangement Breakpoint Database (GRaBD; http://archive.uwcm.ac.uk/uwcm/mg/grabd/) and the European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations (ECARUCA; http://agserver01.azn.nl:8080/ecaruca/ecaruca.jsp). The aim of these databases is to co-ordinate research into rare phenotypes and their underlying rearrangements with the goal of pooling research and possibly identifying the causative genes involved such as in the 17q21.3 microdeletion syndrome (Shaw-Smith et al. 2006).

1.2.5 Normal variation within the human genome

The publication of the human genome reference sequence by the Human Genome Project provided a huge resource with which to compare the genomes

of patients with clinical abnormalities. However, differences are seen between the genomes of apparently normal individuals and it is believed that this heterogeneity within the human genome might partly account for the huge amount of phenotypic diversity seen between individuals. Two initial studies into this variation using a 1Mb large insert clone microarray and an oligonucleotide microarray comparing 55 and 20 individuals respectively, revealed variations ranging from 100Kb to 2Mb often encompassing known genes (Iafrate et al. 2004; Sebat et al. 2004). A further study using the end sequence reads from a fosmid library created from an alternative individual to the human genome reference sequence source revealed 297 structural variation sites greater than 8Kb (Tuzun et al. 2005). A more recent study into this copy number variation (CNV) using 270 individuals from 4 different populations showed that 12% of the genome was identified as being polymorphic (Redon et al. 2006). The data produced by this study has been incorporated into the UCSC and Ensembl web browsers, providing an invaluable tool for the comparison of patient data with known polymorphisms.

CNV data is vital in the study of the effect of chromosomal rearrangements on phenotype. Once a rearrangement has been discovered, it must be determined whether it is causal to the phenotype or is a polymorphism with no phenotypic effect. Analysis of the CNV data for a particular genomic region will identify common polymorphisms with no phenotypic effect. In addition, analysis of a proband's parental DNA will establish whether the event is *de novo* or familial and likely to be causal to the aberrant phenotype.

Whilst a CNV itself might not be instrumental in directly causing a disease phenotype, it might predispose. One study of 30 patients with Thrombocytopenia-Absent Radius (TAR) Syndrome showed that all patients had a deletion at 1q21.1 with a minimal critical region of approximately 200Kb containing 12 genes (Klopocki et al. 2007). Only 25% of these deletions were found to be *de novo*,

with the remaining deletions inherited maternally or paternally from phenotypically normal parents. These results led to the hypothesis that although the syndrome was associated with the deletion, an additional event was required for presentation of the phenotype, suggesting that the region at 1q21.1 is a susceptibility locus.

1.3 Mechanisms behind chromosomal rearrangements

Cellular DNA continuously undergoes a process of breakage and repair which can lead to chromosome rearrangements after the mis-repair of double strand breaks.

1.3.1 Double strand breaks

Double strand breaks can occur along DNA as a result of environmental factors such as ultra-violet and ionising radiation or as a result of normal metabolic function such as degradation by the topoisomerase type II protein, Spo11. Spo11 uses its catalytic tyrosine to attack the phosphodiester backbone of DNA, creating a covalent bond between itself and the 5’ end of the break.

After formation of double strand breaks in the DNA a cell is able to maintain its integrity by either of 2 cellular processes which occur during the cycle of a normal cell; Homologous recombination and non-homologous end joining. Homologous recombination is a high fidelity mechanism which typically occurs without any alteration in the genetic sequence as it relies on a region of almost identical sequence to act as template. However, non-homologous end joining is often accompanied by the loss or addition of nucleotides as the free strands of DNA are altered during the repair of the DNA strands.

Evidence suggests that the mechanism behind the repair of double strand breaks in recurrent rearrangements is homologous recombination as the breakpoints for

these genomic disorders tend to fall within repeat regions. For non-recurrent rearrangements, no mechanism has so far been definitively identified.

1.3.2 Homologous recombination

Recombination was first observed in *D. melanogaster* where it was seen that blocks of genes from homologous chromosomes could be exchanged during a crossing over event generating increased genetic diversity (Morgan 1910). Current models of homologous recombination rely on two sequences with a high degree of homology aligning, followed by the formation of a double strand break and the subsequent re-forming of the phosphodiester bonds to form uninterrupted DNA strands.

Recombination is thought to occur via the formation of an intermediate structure known as the Holliday Junction (Holliday 1964). There are currently 2 proposed methods for the formation of this structure and 2 alternatives for its resolution.

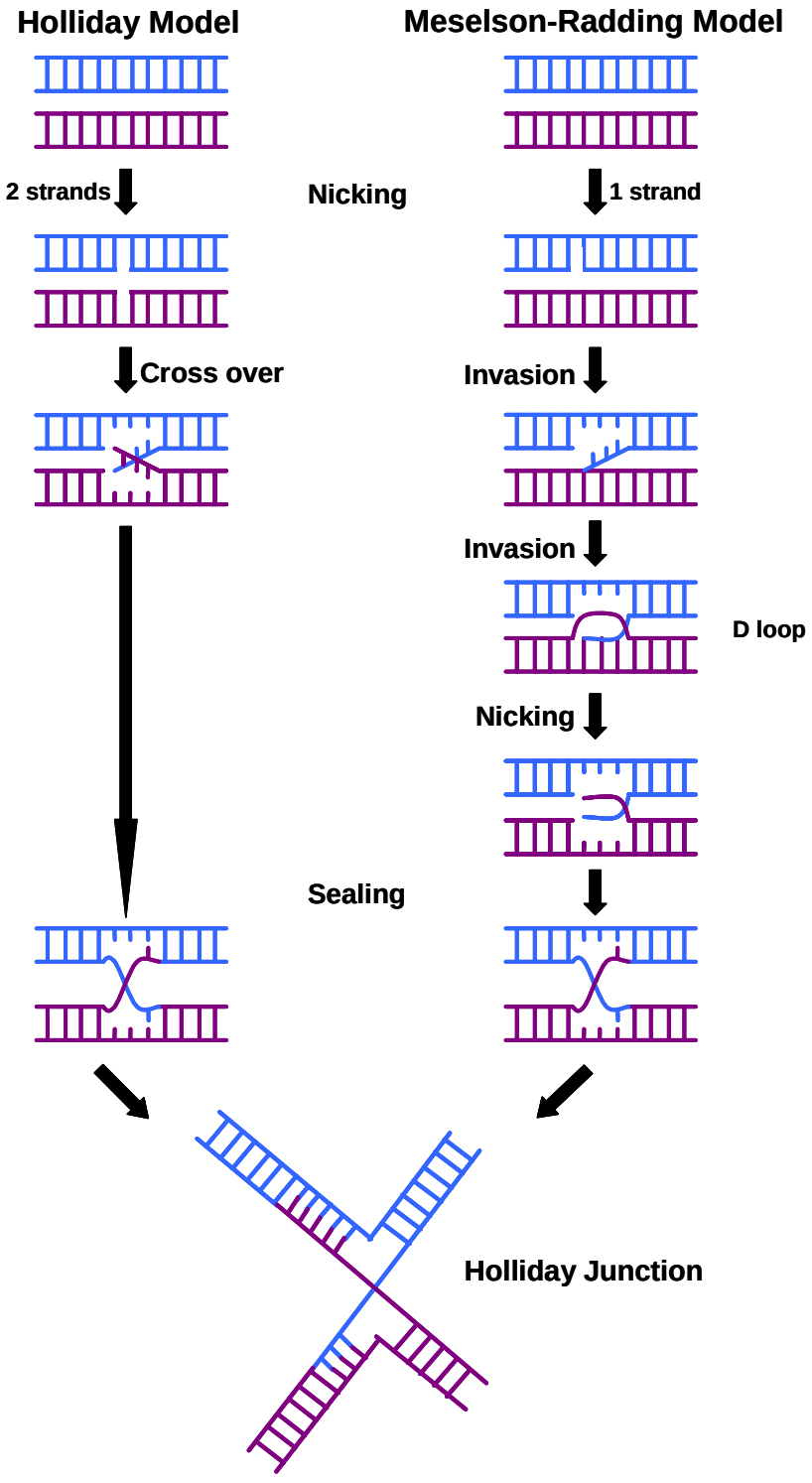


Figure 1.8 Comparison between the Holliday and Meselson-Radding models for formation of the Holliday Junction intermediate structure during recombination.

For the formation of the Holliday junction, the first method proposes that two homologous sequences are aligned and a nick is made in one strand of each promoting strand exchange which occurs at the site of the nicks (Holliday 1964). In the second model, only one nick is needed creating a 5' phosphate end which invades the homologous sequence and causes strand displacement (creating a D-loop) (Meselson and Radding 1975). Both models are compared in Figure 1.8.

The length of homologous sequence present has been shown to have a big impact on recombination within *E.coli* and yeast and regions of mis-match within the homologous regions also have an effect (Watt et al. 1985; Hua et al. 1997). Homologous recombination in somatic mammalian cells is believed to require a minimum of 134-232bp of uninterrupted homology (Waldman and Liskay 1988) and meiotic homologous recombination in humans is believed to require a minimum of 337–456bp (Reiter et al. 1998).

For the resolution of the Holliday junction, the first method proposes that all four strands are cut at the crossover site resulting in recombinant chromosomes. The second method proposes that the Holliday structure rotates at the crossover site and two strands are cut (if the original un-nicked strands are cut, recombinants are formed but if the original nicked strands are cut, recombinant chromosomes are not formed) (Figure 1.9).

Variations in the rates of recombination within the human genome have revealed hotspots of recombination although the exact reason for this remains unknown (Crawford et al. 2004; McVean et al. 2004). It is estimated that approximately 80% of recombination occurs within 10-20% of the sequence with an average of one hotspot every 50Kb along the genome (Myers et al. 2005). Known regions of meiotic recombination hotspots have been collated from the literature in an effort

to improve our understanding of the mechanisms behind this phenomenon (Nishant et al. 2006).

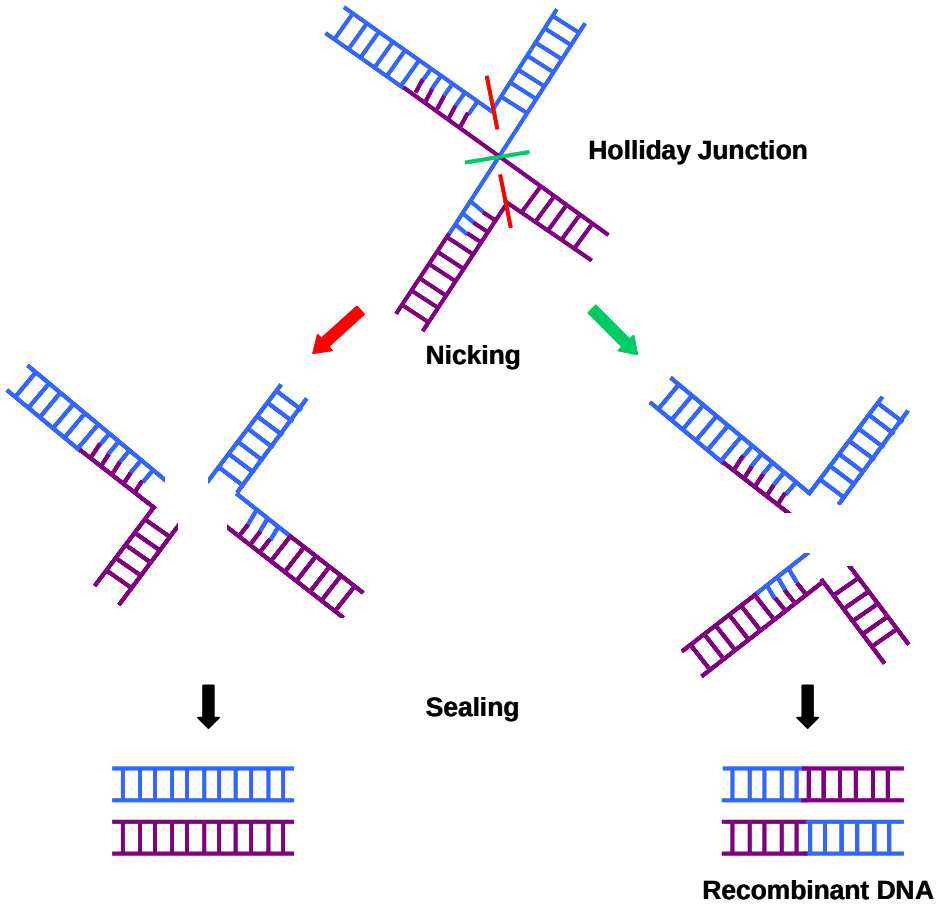


Figure 1.9 Schematic showing the resolution of the Holliday junction during the process of homologous recombination. If the nicking occurs in the previously unaffected strands and no strand invasion has occurred (not depicted on diagram), no recombinant DNA is observed (red). If the 2 strands that were nicked during formation of the Holliday junction are nicked again, the result is recombinant DNA (green).

1.3.3 Non-homologous end joining

When double strand breaks occur in non repetitive regions of the genome and a region of homology cannot be found to act as template for the repair, then the cell relies on non-homologous end joining. Typically the broken ends will be

incompatible and the rejoining will require nucleases to remove nucleotides and/or polymerases to fill in nucleotides, resulting in the characteristic gain or loss of nucleotides at the repaired junctions.

1.3.4 Genome architecture associated with rearrangements

Certain genetic diseases are the result of recurrent chromosomal rearrangements at particular regions of the genome. Analysis of these rearrangements and use of the human genome reference sequence across these regions has identified certain features that may be associated ranging from recombinogenic sequence motifs and repeat sequences to secondary DNA structures that leave the DNA prone to double strand breaks.

1.3.4.1 Chi sequences

The Chi element is known to be a mediator of prokaryotic recombination. The RecBCD enzyme in E.coli recognises the Chi sequence motif GCTGGTGG and binds, promoting recombination (Smith et al. 1981). In humans a Chi-like sequence has been observed with the consensus; GC[A/T]GG[A/T]GG (Krowczynska et al. 1990).

1.3.4.2 Translin associated motifs

Translin motifs have been associated with the translocation breakpoints in many lymphoid malignancies. The consensus sequence observed (ATGCAG and GCCC[A/T][G/C][G/C][A/T]) are identified by the translin protein which binds to the DNA and is believed to be involved in chromosomal translocations (Aoki et al. 1995).

1.3.4.3 Topoisomerase I and II sites

Topoisomerases catalyse DNA unwinding. Topoisomerase I cuts a single strand whilst Topoisomerase II cuts both strands. Topoisomerase II sites also function as chromosome scaffold attachment sites – a chromosome configuration that brings distant DNA sequences into close proximity. The Topoisomerase I

consensus sequence is [A/T][G/C][A/T]T (Been et al. 1984) and there are 3 Topoisomerase II recognition sequences;

Vertebrate [A/G]N[T/C]NNCNNG[T/C]NG[G/T]TN[T/C]N[T/C]

Drosophila GTN[T/A]A[C/T]ATTNATNNG

Invertebrate [T/C][A/C]CNTAC[C/G][C/T]CC[T/G][T/C][T/C]TNNC

(Sander and Hsieh 1985; Spitzner and Muller 1988; Kas and Laemmli 1992).

1.3.4.4 Immunoglobulin heptamers

The immunoglobulin gene locus undergoes recombination which results in variation within the antibody response system. The immunoglobulin recognition sequence consists of a heptamer with the consensus GATAGTG or CACAGTC separated from a nonamer with the consensus sequence ACAAAAACC (Sakano et al. 1979; Chen et al. 1989).

1.3.4.5 Mini satellite sequences

The discovery of hypervariable mini satellite DNA sequences and their involvement in human polymorphisms lead to the suggestion that they might be involved in chromosomal rearrangements (Wyman and White 1980). They are comprised of short tandemly repeated DNA fragments with a core sequence of GGGCAGGC[A/G]G (Jeffreys et al. 1985; Jeffreys 1987).

1.3.4.6 Purine/pyrimidine tracts

Purine (A/G) and pyrimidine (T/C) tracts have been observed to be recombinogenic (Boehm et al. 1989; Majewski and Ott 2000) and a study of 96 chromosome translocations revealed that around breakpoints alternating purine/pyrimidine tracts (2-30bp) were underrepresented, polypurine tracts (2-23bp) were overrepresented and polypyrimidine tracts (2-44bp) were overrepresented (Abeysinghe et al. 2003).

1.3.4.7 SINE elements

Short Interspersed Nuclear Elements (SINE elements) are transposable repeat elements which are typically 100-400bp long (Smit 1996). There are 2 major classes; Alu repeats and the MIR family (Mammalian-wide interspersed repeat).

The Alu repeat is the most abundant sequence in the human genome with an estimated 1 million copies accounting for just over 10% (Lander et al. 2001). The full Alu repeat is approx 280bp long flanked by direct repeats of 6-18bp. Alu repeats have a relatively high GC content (56%) and are reported to preferentially be found in the pale bands of G-banded chromosomes (Korenberg and Rykowski 1988).

MIR repeats have approximately 450,000 copies comprising 2.5% of the human genome (Lander et al. 2001).

1.3.4.8 LINE elements

Long interspersed nuclear elements (LINE elements) are typically 6-8Kb in length and the estimated 868,000 copies present in the human genome are responsible for approximately 21% of its composition (Lander et al. 2001). There are 3 classes of LINE elements; L1, L2 and L3. They are a type of transposable element encoding a reverse transcriptase and a DNA-nick-looping enzyme, allowing them to move about the genome autonomously.

1.3.4.9 Long terminal repeats

Long terminal repeats (LTRs) account for approximately 8% of the human genome (Lander et al. 2001). Many LTRs contain viral promoter, enhancer and polyadenylation signals and have functional consequences (Landry and Mager 2002).

1.3.4.10 Class II transposons

Class II Transposons move by excision and reintegrate into the genome without using an RNA intermediate. They are characterized by terminal inverted repeats of 10-500bp in length.

1.3.4.11 Palindromes

Palindromic DNA can lead to unstable single stranded hairpin or double stranded cruciform structures which are prone to double strand breaks. The DNA then illegitimately recombines to form the translocated chromosomes. In particular, the motif TTTAAA has been shown to bend DNA at sites of recombination (Singh et al. 1997).

1.3.4.12 Segmental duplications

Segmental duplications or Low Copy Repeats (LCRs) are regions of the genome that have been duplicated and range from 1-200Kb in length (Lander et al. 2001). These regions of homologous sequence have been implicated in the aetiology of several genomic disorders such as Sotos syndrome (Kurotaki et al. 2005; Visser et al. 2005), Charcot-Marie-Tooth disease type 1A and hereditary neuropathy with liability to pressure palsies (Shaw et al. 2004) and Smith-Magenis syndrome and dup(17)(p11.2p11.2) syndrome (Smith et al. 1986; Potocki et al. 2000). The inter- and intra- chromosomal segmental duplications within the human genome greater that 10Kb are summarised in Figure 1.10.

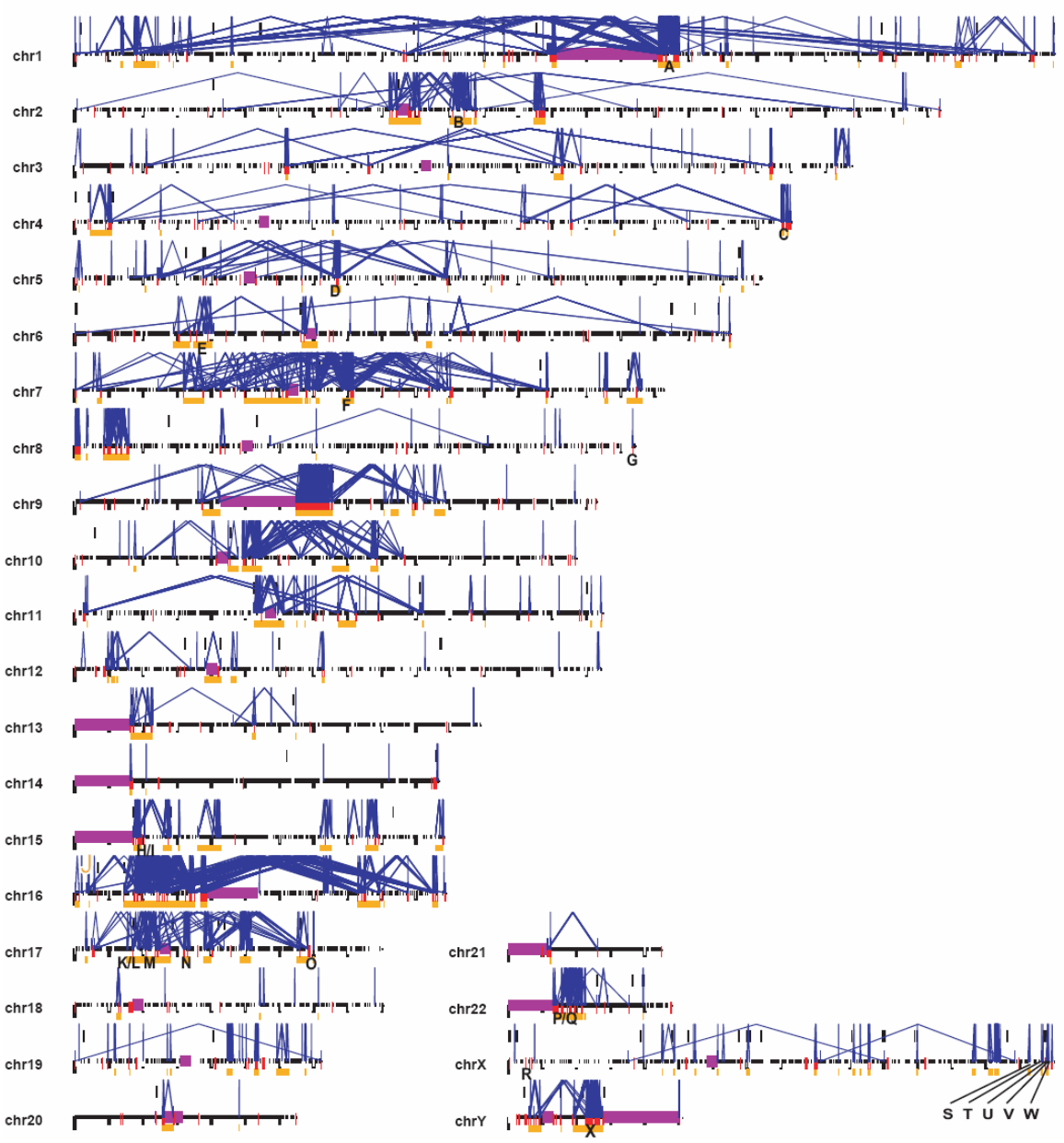


Figure 1.10 ***Patterns of interchromosomal (red) and intrachromosomal (blue) segmental duplications of ≥10Kb with ≥95% homology across the human genome. Colour coding; Purple - Areas not sequenced as part of the human genome project (Acrocentric chromosome arms, heterochromatic regions, centromeres) and orange - unique regions of the genome 50Kb-10Mb in size encompassed by intrachromosomal duplications. 24 regions labelled A to X are known hotspots for genomic rearrangement (Full details in Appendix A2) (Bailey et al. 2002).***

1.4 Viral Integration into the human genome

Human chromosomes are also susceptible to rearrangement by integration from viral genomes. Viruses have the ability to invade human cells, particularly in immunocompromised patients, and have been seen to integrate into the human genome. A well characterised example of targeted viral integration is the adeno-associated virus (AAV). The virus has 2 terminal repeats of 145bp and evidence suggests that viral insertions are in a tandem head to tail orientation via these repeats (Cheung et al. 1980). More recently, this integration has been shown to be a targeted event with integration into the distal portion of chromosome 19q (Samulski et al. 1991). Generation of sequence across and subsequent analysis of the viral-cellular junctions revealed a 2-3bp homology between the viral and cellular DNAs.

1.4.1 The Human Herpes Virus

The human herpes virus (HHV) is ubiquitous in the human population and normally exists as an extra chromosomal element in the host (Kondo et al. 1991; Luppi et al. 1999). There are 8 members in the human herpes virus family; HHV-1 to HHV-8. HHV-6 is a lymphotropic virus which has been identified as a causative agent in the illness exanthum subitum, also known as Roseola infantum, whose patients present with rash, fever and febrile convulsions (Kondo et al. 1991). The data suggested that the viral DNA may persist within cells in a latent form before being reactivated. HHV-6 exists as two variants; HHV-6A and HHV-6B which are highly conserved genetically with the amino acid sequence of genes sharing 75-95% homology (Lindquester et al. 1996).

HHV-6 was initially isolated from peripheral blood leukocytes in 6 patients with lymphoproliferative disorders and 4 healthy donor samples (Salahuddin et al. 1986). Subsequent investigations isolated HHV-6 from HIV-infected patients (Downing et al. 1987; Tedder et al. 1987; Lopez et al. 1988), immunocompromised patients following organ transplantation (Ward et al. 1989)

and from healthy individuals (Pietroboni et al. 1988; Saxinger et al. 1988; Okuno et al. 1989; Harnett et al. 1990). The human herpes virus has the ability to cause disease in immuno-compromised patients, such as transplant recipients and is thought to be a cofactor in the progression of HIV disease (Salahuddin et al. 1986). Studies have shown that up to approximately 95% of the global population have been infected (Aberle et al. 1996) with infection usually occurring early in life (Briggs et al. 1988; Knowles and Gardner 1988).

1.4.2 Evidence of integration and inheritance

A study of 3 patients using Pulsed field gel electrophoresis and Southern blot analysis showed that the virus was attached to high molecular weight DNA and the authors hypothesised that this DNA was human chromosomal DNA (Luppi et al. 1993). Further investigations into these 3 patients by FISH showed that the virus integrated into 17p13 in all 3 cases (Torelli et al. 1995), later confirmed to be close to or within the telomeric sequence of 17p (Morris et al. 1999).

A further study into the prevalence of HHV-6 in healthy individuals using quantitative PCR revealed that 36% (9 out of 25) of the volunteers studied exhibited a viral burden in their blood (Clark et al. 1996). One of these volunteers showed a consistently high burden over the 10 months of the study, a suggested result of integration into the human genome leading to the conclusion that the HHV-6 genome has integrated into the genome of approximately 3% of the British population.

More recently evidence of inheritance of the integrated viral genome has been obtained. In one study, FISH identified integration of the HHV-6 virus into 1q44 of a female patient. Subsequent analysis of the patient's offspring and grandchildren revealed that her son and granddaughter both showed HHV-6 integration at the same 1q44 locus (Daibata et al. 1998). An additional study

identified a family in which the mother carried HHV-6 integration at 22q13, the father at 1q44, and their daughter at both locations (Daibata et al. 1999).

1.4.3 The structure of the HHV-6B Virus

Two strains of the HHV-6B genome have been fully sequenced, Z29 (NCBI accession number; NC_000898) and HST (NCBI accession number; AB021506). The Z29 strain was fully characterised by restriction digest patterns using BamHI, ClaI, HindIII and SalI (Pellett et al. 1990). This data was subsequently verified and the restriction digest fragments cloned to provide a valuable resource for investigations into HHV-6B infection (Lindquester et al. 1996). The HHV-6B virus is a linear genome of approximately 160Kb consisting of a central segment of approximately 141Kb of unique sequence flanked by two direct repeats (termed DR_L and DR_R) which can vary in length from 10 to 13Kb (Figure 1.11). These repeats containing 2 smaller blocks of sequence $(GGGTTA)_n$ similar to that found at the telomeres of human chromosomes, $(TTAGGG)_n$ (Moyzis et al. 1988; Thomson et al. 1994) and it is postulated that this might be the mechanism by which the virus is able to integrate into the human genome.

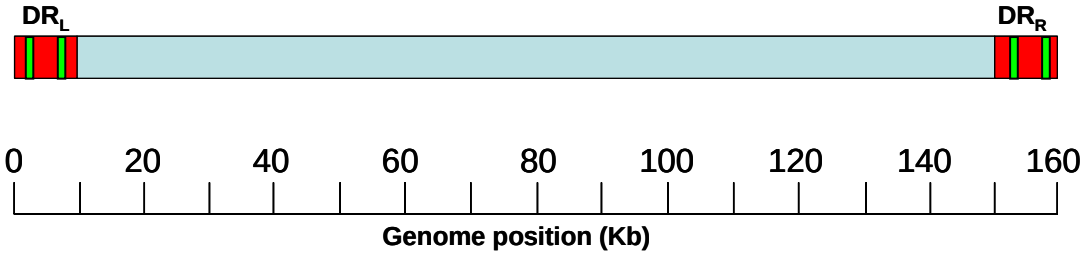


Figure 1.11 Schematic of the HHV-6 genome showing unique sequence (blue) of the viral gnome flanked by direct repeats (red) at the left and right ends of the genome (DR_L and DR_R). These direct repeats each contain two regions of repeats (green) similar to the human telomeric repeat sequence.

1.5 Aims of this thesis

1.5.1 Use and adaptation of current techniques for the mapping of rearrangement breakpoints in three patients with abnormal phenotypes

The aim of this study is to map and sequence 3 apparently balanced reciprocal translocations in patients with learning difficulties and/or physical developmental delay with a goal of identifying the underlying cause of the phenotype and the mechanism mediating the rearrangements. To do this I have implemented molecular cytogenetic tools such as FISH and genomic microarrays combined with flow sorting technologies (array painting) followed by PCR to amplify junction fragments.

1.5.2 Development of techniques for the investigation of rearrangement breakpoints

As more translocations are mapped, it has become increasingly apparent that the sequence surrounding the breakpoints may harbour previously undetected abnormalities (Kumar et al. 1998; Astbury et al. 2004; Patsalis et al. 2004; Ciccone et al. 2005; Gribble et al. 2005). These abnormalities may hamper the use of PCR in the refining of translocation breakpoints and the amplification of breakpoint junction fragments. An alternative method to array painting and STS PCR mapping is the generation of a custom-made library derived from the patient's flow sorted derivative chromosomes and the selection of clones chimeric for both donor chromosome sequence either side of the breakpoint. By creating a fosmid library, approximately 40Kb of sequence around the breakpoint can be obtained and studied for changes compared to the reference sequence. A library approach may also be beneficial to obtain sequence across a breakpoint if other methods fail to sufficiently refine it. To facilitate the mapping of breakpoints I will develop a custom-made fosmid library approach taking advantage of chromosome sorting technology to enrich the library for derivative chromosome material. Once established, the benefits of a library approach can be compared

to methods such as FISH, PCR and microarrays, including commercially available tools.

1.5.3 Bioinformatic analysis of the genomic architecture surrounding genomic rearrangement breakpoints

The generation of sequence across constitutional rearrangement breakpoints and analysis of its composition may reveal clues as to the mechanism behind the rearrangement. Genome browsers and web-based tools will be used to search for structures known to be involved in other rearrangements and to search for motifs common to the breakpoints and compare this data with the analysis performed on the published constitutional translocation breakpoints listed in Table 1.1.

1.5.4 Application of techniques developed for the mapping of rearrangement breakpoints to the investigation of viral integration in the human genome

In this project I aim to utilise currently available tools and develop and apply new methods for the investigation of translocation breakpoints to the investigation into viral integration sites (Figure 1.12). I plan to characterise the integration site in 3 patients believed to have integrated Human Herpes Virus-6 genomes. I will investigate the chromosomal location of the viral integration using FISH and whether the complete virus is integrated using comparative genomic hybridisation microarray analysis. By generating a custom-made fosmid library and identifying clones that contain the integration sites, information about the sequence at the integration sites can be obtained, perhaps giving an indication as to the mechanism by which the virus integrates.

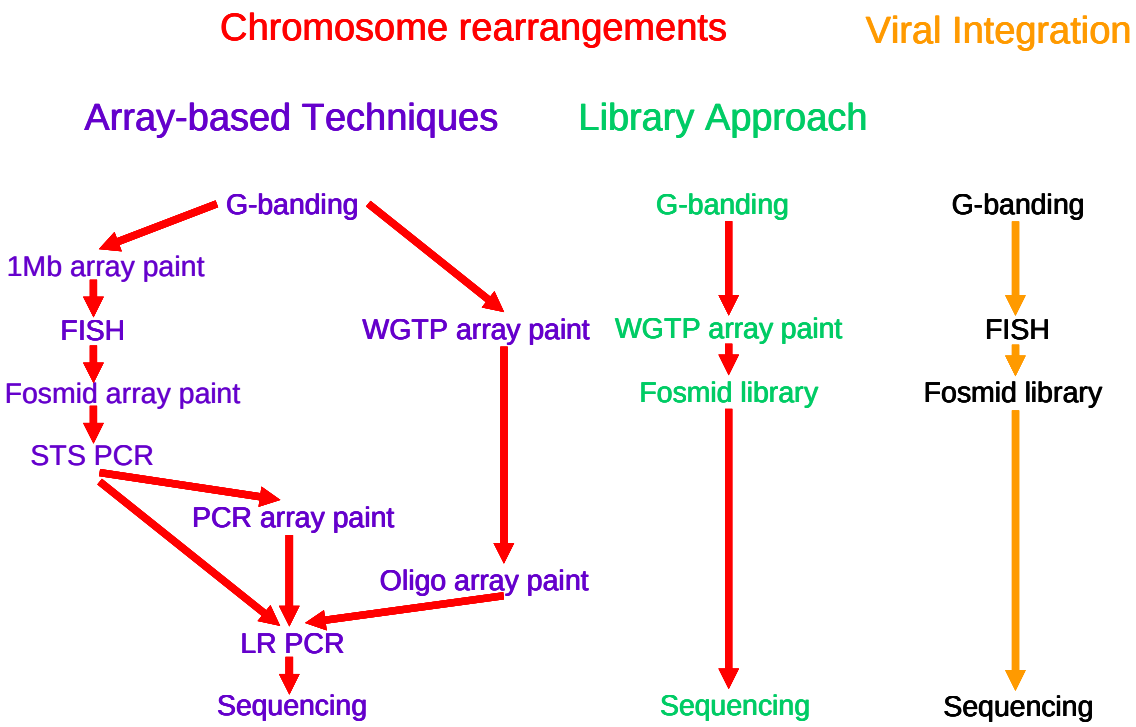


Figure 1.12 Summary of the proposed techniques for the analysis of rearrangement breakpoints and viral integration sites in the human genome.

2 Materials and methods

2.1 Cell culture

Lymphoblastoid cell lines were established from blood samples for patients t(2;7)(q37.3;p15.1), t(3;11)(q21;q12) and t(7;13)(q31.1;q21.3) by the European Cell and Culture Collection (ECCAC), Porton Down and the Cell Bank at Wessex Regional Genetics Laboratory (WRGL) for patient t(2;6)(q21.1;q25.1).

For patients AMD, AS and KK showing high viral load, lymphoblastoid cell lines were established from peripheral blood mononuclear cells (PBMCs) at the Department of Virology, Royal Free and University College Medical School of UCL, London.

All lymphoblastoid cell lines were cultured at 37°C in RPMI 1640 media (Invitrogen) supplemented with 16% Foetal Calf Serum (FCS; Invitrogen), 2mM L-glutamine/ 100U/ml penicillin/ 100µg/ml streptomycin (Sigma).

2.2 Isolation of Patient DNA

DNA was isolated from the patient derived cell lines by various methods for different protocols. Whole genomic DNA was used for array CGH and LR PCR. Flow sorted derivative chromosomes were used directly for the generation of fosmid libraries and for array painting onto WGTP arrays and custom-made oligonucleotide microarrays (NimbleGen Systems Inc.). After amplification, flow sorted derivative chromosomes were also used for STS PCR mapping and array painting onto fosmid microarrays.

2.2.1 Genomic DNA extraction

Whole genomic DNA was extracted using the Qiagen Blood and Cell Culture DNA Midi kit according to the manufacturer's protocol. Briefly, $2x10^7$ cells were centrifuged at 1500g for 10min. The pellet was washed twice with Phosphate

Buffered Saline (PBS; 1.37M NaCl/ 27mM KCl/ 100mM Phosphate buffer) and resuspended in 2ml ice-cold (4°C) PBS. 2ml ice-cold buffer C1 (Cell Lysis Buffer; 1.28M sucrose/ 40mM Tris-Cl pH7.5/ 20mM $MgCl_2$/ 4% Triton X-100) and 6ml ice-cold water were added to the cells, inverted and incubated on ice for 10min. The cells were centrifuged at 1300g for 15min and resuspended in 1ml ice-cold buffer C1 and 3ml ice-cold water before being recentrifuged at 1300g for 15min. 5ml buffer G2 (General Lysis Buffer; 800mM Guanidine HCl/ 30mM Tris-Cl pH8/ 30mM EDTA pH8/ 5% Tween-20/ 0.5% Triton X-100) was added and the nuclei resuspended by vortexing. 95µl 20mg/ml Qiagen protease was added prior to incubation at 50°C for 1hr. 100/G Genomic tips were prepared by allowing 4ml Buffer QBT (Equilibration Buffer; 750mM NaCl/ 50mM 3-[N-Morpholino] propanesulphonic acid (MOPS) pH7/ 15% isopropanol/ 0.15% Triton X-100) to flow through prior to addition of the nuclei. The tip with sample was then washed twice with 7.5ml Buffer QC (Wash Buffer; 1M NaCl/ 50mM MOPS pH7/ 15% isopropanol). Genomic DNA was eluted with 5ml Buffer QF (Elution Buffer; 1.25M NaCl/ 50mM Tris-Cl pH8.5/ 15% isopropanol). The DNA was precipitated by the addition of 3.5ml isopropanol and inversion of the collection tube. DNA was spooled using a sterile loop, transferred to a tube containing 50µl water and allowed to resuspend overnight at 4°C. The DNA concentration was measured using a TD-360 fluorometer (Turner Designs) and a 0.6% agarose gel with ethidium bromide was run to assess yield and quality.

2.2.2 Flow sorting of derivative chromosomes (Bee Ling Ng – Team 70)

Chromosomes from cell lines were prepared in polyamine buffer (pH7.5; 15mM Tris/ 2mM EDTA/ 0.5mM Ethylene glycol-bis N,N,N'N'-tetra acetic acid (EGTA; Sigma)/ 80mM KCl/ 3mM dithiothreitol/ 0.25% Triton X-100 (Sigma)/ 0.2mM spermine (Sigma)/ 0.5mM spermidine (Sigma)/ 20mM NaCl) and stained with 5ug/ml of Hoechst (Sigma) and 40ug/ml of Chromomycin A3 (Sigma). The stained chromosome sample was analysed and flow sorted using a MoFlo high speed cell sorter (Dako). 500 chromosomes for DOP PCR were sorted in 33µl

UV treated water and 250,000 chromosomes for array painting and fosmid library procedures were sorted in 250µl sheath buffer (10mM Tris-HCl (pH8.0)/ 1mM EDTA/ 100mM NaCl/ 0.5mM sodium azide).

"Bulk sorted" chromosomes (tubes of 250,000) were subjected to an overnight incubation at 42°C with 15µl 0.25M EDTA/ 10% sodium lauroyl sarcosine (Sigma) and 2µl 20mg/ml Proteinase K (Sigma). Chromosomes were then incubated at room temperature for 40min with 2µl 4mg/ml phenylmethanesulfonyl fluoride (PMSF; Sigma) in 96% ethanol to inactivate the proteinase K. Finally, chromosomes were precipitated with 10µl 5M NaCl, 2µl pellet paint (Novagen) and 770µl 100% ethanol at -20°C overnight.

2.3 Amplification of template DNA

2.3.1.1 REPLI-g

Flow sorted derivative chromosomes were amplified by Repli-G, a commercially available kit (QIAGEN) for whole genome amplification by Multiple Displacement Amplification (MDA). Replication is obtained using a DNA Polymerase capable of synthesising 100Kb of sequence at 30°C.

Tubes of 250,000 flow sorted derivative chromosomes were precipitated and resuspended in 20µl T0.1E (10mM Tris, 0.1mM EDTA) with 1µl of this used for the REPLI-g reaction as per the manufacturer's instructions. Briefly, 1µl of flow sorted chromosomes were added to 1.5µl TE (10mM Tris/ 1mM EDTA) and 2.5µl Buffer D1 (1 part Solution A (0.4M KOH/ 12.5mM EDTA)/ 7 parts nuclease free water) and incubated at room temperature for 3min. The reaction was stopped by the addition of 5µl Buffer N1 (1 part Solution B/ 9 parts nuclease free water). 27µl nuclease-free water, 12.5µl 4x REPLI-g buffer and 0.5µl REPLI-g DNA polymerase were added and the samples incubated at 30°C for 8hr and inactivated at 65°C for 3min.

2.3.1.2 DOP PCR

Flow sorted derivative chromosomes were amplified by degenerate oligonucleotide primer (DOP) PCR. The primers contained 6 bases of degenerate sequence and this coupled with initial low annealing temperatures allowed amplification at multiple sites along the template sequence.

Tubes of 500 flow sorted derivative chromosomes were amplified in 50µl reactions by DOP PCR. 5µl 10xTAPS2 buffer (250mM N-Tris (hydroxymethyl) methyl-3-amino-proane sulphonic acid pH9.3 (Sigma)/ 500mM KCl/ 20mM $MgCl_2$/ 10mM dithiothreitol/ 0.7% ß-mercaptoethanol/ 0.165% Bovine Serum Albumin), 4µl 2.5mM each dNTPs, 0.5µl 5U/µl Amplitaq (PerkinElmer), 2.5µl 1.25% W1 detergent (Sigma) and 5µl 20µM DOP primer were added to 500 chromosomes in 33µl water. For a negative control, 33µl of UV treated water was used. The sequence of the primers used is detailed in Table 2.1.

Primer name	Primer sequence
6MW	CCGACTCGAGNNNNNNATGTGG
DOP1	CCGACTCGAGNNNNNNCTAGAA
DOP2	CCGACTCGAGNNNNNNTAGGAG
DOP3	CCGACTCGAGNNNNNNTTCTAG

Table 2.1 DOP primers used for the amplification of chromosomes and clones by DOP PCR.

Samples were heated at 94°C for 9min followed by 10 cycles of 94°C for 1min, 30°C for 1min 30sec ramping at 0.23°C per second, 72°C for 3min followed by 30 cycles of 94°C for 1min, 62°C for 1min, 72°C for 1min 30sec followed by 72°C for 9min.

2.4 Construction of microarrays

2.4.1 Genomic clone microarrays (fosmid and plasmid microarrays)

2.4.1.1 Clone selection

For fosmid microarray production, fosmid clones were selected at full redundancy from Build35 of the UCSC genome browser (http://genome.ucsc.edu/) to cover the spanning BAC clone and approximately 100Kb of sequence both proximal and distal to this region. In addition, BAC clones were selected at 10Mb spacing from the ”Golden Path”. Fosmid and BAC clones were picked from libraries held within the Sanger Institute.

Plasmid clones were selected to provide maximum coverage of the HHV-6 genome and were kindly supplied by the Centre of Disease Control, Atlanta.

2.4.1.2 Picking

Clones were picked into LB media with appropriate antibiotic (12.5µg/ml chloramphenicol for fosmids, 20µg/ml chloramphenicol for BACs or 100µg/ml ampicillin for plasmids) and incubated at 37°C overnight. All clones were tested for T4 bacteriophage λ and *Pseudomonas aeruginosa* contamination before being sent for prepping.

2.4.1.3 DNA prepping (Carol Carder – Team 63)

Clones were grown overnight in 2ml 96-well boxes (Beckman) at 37°C with shaking in 2xTY media with the appropriate antibiotic (12.5µg/ml chloramphenicol for fosmids, 20µg/ml chloramphenicol for BACs or 100µg/ml ampicillin for plasmids). 250µl of the overnight culture was transferred into a 96 well plate (Costar), centrifuged for 2min at 800g at 20°C and the pellets resuspended in 25µl Solution 1 (50mM Glucose/ 10mM EDTA/ 5mM Tris pH8.0). 25µl Solution 2 (0.2mM NaOH/ 1% SDS) was added and the plates incubated at room temperature for 5min followed by the addition of 25µl 3M KOAc pH5.5 and a

further 5min incubation at room temperature. The contents of each well were transferred to a 96 well filter plate (Millipore), centrifuged for 2min at 800g at 20°C into 100µl isopropanol and incubated at room temperature for 30min. The plates were spun at 1300g for 20min at 20°C and the pellets air dried for 15min. The pellets were washed twice with 70% ethanol and air dried before being resuspended in 5µl T0.1E (10mM Tris, 0.1mM EDTA) with 10µg/ml RNase (Sigma).

2.4.1.4 DOP PCR amplification of plasmid, fosmid and BAC clones

All clones were amplified using 3 separate DOP primers (DOP 1, DOP 2 and DOP 3 in Table 2.1). 50µl DOP PCR reactions were performed in 96 well plates (Costar). 28µl water, 5µl 10xTAPS2 buffer (250mM N-Tris (hydroxymethyl) methyl-3-amino-proane sulphonic acid pH9.3 (Sigma)/ 166mM $(NH4)_2SO_4$/ 25mM $MgCl_2$/ 0.7% ß-mercaptoethanol/ 0.165% Bovine Serum Albumin), 5µl 20µM DOP primer, 4µl 2.5mM each dNTPs, 2.5µl 1% W1 detergent (Sigma) and 0.5µl 5U/µl Amplitaq (PerkinElmer) were added to 5µl of clone DNA (section 2.4.1.3) and subjected to temperature cycling on an MJ-Thermocycler; 94°C for 3min followed by 10 cycles of 94°C for 1min 30sec, 30°C for 2min 30sec ramping at 0.1°C per sec to 72°C, 72°C for 3min followed by 30 cycles of 94°C for 1min, 62°C for 1min 30sec, 72°C for 2min and finally 72°C for 8min. 5µl of each reaction was analysed using 2.5% agarose gel electrophoresis with ethidium bromide staining (Figure 2.1).

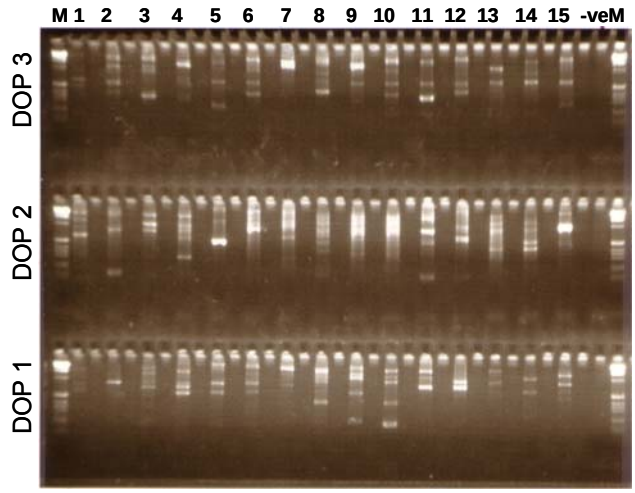


Figure 2.1 PCR amplification of 15 plasmid clones using primers DOP 1, DOP 2 and DOP 3 run on 2.5% agarose gel run with 1Kb marker. Unlabelled wells were left unloaded.

2.4.1.5 Amino-linking of DOP PCR products

Prior to amino-linking PCR, all 3 DOP PCR reactions for each clone were combined. 3µl of this combined product was used as template in an amino-linking PCR to attach an amino group to the 5' end of products to allow for covalent bonding of the product to the glass microarray slides.

90µl amino-linking PCR reactions were performed using 3µl of DOP template, 63.6µl water, 9µl 10x Amino-linking buffer (500mM KCl/ 25mM MgCl2/ 50mM Tris pH8.5), 9µl 2.5mM each dNTPs, 4.5µl 200ng/µl Amino-primer (GGAAACAGCCCGACTCGAG with an amino group attached to the 5' end) and 0.5µl 5U/µl Amplitaq (PerkinElmer). Reactions were performed on an MJ-Thermocycler; 95°C for 10min followed by 35 cycles of 95°C for 1min, 60°C for 1min 30sec, 72°C for 7min followed by 72°C for 10min. 2µl of each reaction was analysed using 2.5% agarose gel electrophoresis with ethidium bromide staining (Figure 2.2).

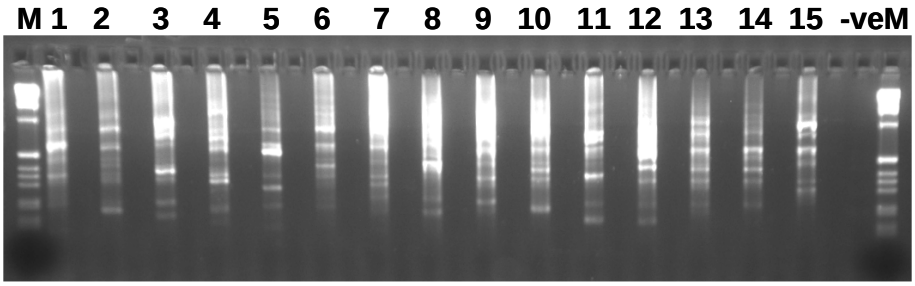


Figure 2.2 Amino-linking PCR for 15 plasmid clones run on 2.5% agarose gel with 1Kb marker. Unlabelled wells were left unloaded.

2.4.1.6 Spotting onto slides (Cordelia Langford & Team 77)

45µl PCR products (section 2.4.1.5) were added to 15µl 4x spotting buffer (1M sodium phosphate buffer pH8.5/ 0.001% N-lauroyl-sarcosine/ 0.4% azide) in multiscreen filter plates (Millipore) and filtered by centrifugation at 550g for 10min into 96 well plates (Falcon). Using a Qiagen RapidPlate liquid handling robot, 15µl of each filtered sample was transferred into a Genetix X6004 384-well plate. DNA products were arrayed onto Amine Binding slides (Motorola) at 20-25°C at 40-50% humidity using a MicroGridII array gridding robot (Genomic Solutions) equipped with tungsten spotting pins. The microarray slides were incubated for 24-72hr at room temperature in 70-80% humidity, incubated for 5min in 1% ammonium hydroxide at room temperature followed by a further incubation for 5min in 0.1% sodium dodecyl sulphate at room temperature before being rinsed in Double Distilled Water (DDW) at room temperature followed by immersion in 95°C DDW for 2min. Slides were then transferred to ice-cold DDW and then briefly washed in room temperature DDW before drying by centrifugation at 150g for 5min at room temperature. Microarray slides were stored in a dessicator at room temperature until use.

2.4.2 PCR product arrays

2.4.2.1 PCR primer design (James Morris – Team 117)

18-22bp oligonucleotide primers with 40-60% GC content were designed to repeat-masked sequence to amplify 800-1200bp products tiling the region of interest. Forward primers were synthesised with an amino group attached to their 5' end.

2.4.2.2 PCR amplification of products for microarray production

Primer pairs were used to amplify products from genomic DNA (Promega) for patient t(2;7)(q37.3;p15.1) or viral DNA (kindly supplied by Hoe Nam Leong and Duncan Clarke at UCL, London) for the HHV-6 microarray. All primer pairs were also tested on water as a negative control.

60µl PCR reactions were conducted in 96 well plates (Costar) with 15ng template DNA in 5µl (or 5µl water for negative control), 1.5µl 200ng/µl forward primer, 1.5µl 200ng/µl reverse primer, 6µl 10xPCR Buffer (500mM KCl/ 50mM Tris pH8.5/ 25mM $MgCl_2$), 3µl 10mM each dNTPs, 0.375µl 5U/µl Amplitaq (PerkinElmer) and 42.625µl water. Samples were subjected to temperature cycling of 95°C for 5min followed by 35 cycles of 95°C for 1min, 65°C decreasing by 0.3°C per cycle for 1min 30sec, 72°C for 1min 30sec followed by 72°C for 5min. PCR products were spotted onto glass slides as described in section 2.4.1.6.

2.5 Hybridisation and analysis of microarrays

2.5.1 Hybridisation and analysis of in-house arrays (PCR, fosmid, WGTP)

All microarrays were hybridised on an HS4800 Hybridisation Station (Tecan) using small chambers (51 x 20mm) (fosmid microarrays and PCR product microarrays) or large (63.5 x 21mm) chambers (Whole Genome Tile Path

(WGTP) microarrays). All details given remain the same regardless of which microarray was used unless otherwise stated.

2.5.1.1 Labelling

Template DNA (150ng genomic DNA for CGH, DNA from 250,000 flow sorted derivative chromosomes or 1µl REPLI-g amplified DNA for array painting) was labelled using reagents from the BioPrime Labelling Kit (Invitrogen) with modifications as stated to allow for labelling of the DNA with fluorescent dyes. DNA was mixed with 60µl Random Primers Solution and water to make up to 130.5µl, heated to 100°C for 15min and then cooled on ice prior to the addition of 15µl 10xdNTP mix (1mM dCTP/ 2mM dATP/ 2mM dGTP/ 2mM dTTP), 1.5µl 1mM Cy3 or Cy5 labelled dCTP (NEN Life Science) and 3µl Klenow fragment. Samples were incubated at 37°C overnight and the reactions terminated by the addition of 15µl Stop Buffer. For CGH onto the WGTP microarray, 300µl labelling reactions were set up per sample and per reference in duplicate with the dyes reversed so that dye swap experiments could be performed.

Labelled samples were cleaned up using Microcon YM-30 columns (Millipore); 150µl labelled sample and 150µl water (300µl combined labelling reactions for WGTP array) were applied to the column and spun at 12,000g for 5min, then washed with 300µl water and spun as before. 100µl water was added to the sample and the inverted column spun in a fresh tube at 400g for 2min to collect the sample. 3µl of each sample was analysed using 2.5% agarose gel electrophoresis with ethidium bromide staining to check for successful amplification.

2.5.1.2 Precipitation of samples

For each microarray a sample tube and prehybridisation tube was required; for the sample tube 80µl Cy3 labelled DNA, 80µl Cy5 labelled DNA and 135µl Human Cot1 DNA (Invitrogen) were precipitated with 35µl 3M Sodium Acetate

pH5.2 and 1ml 100% ethanol. For prehybridisation tubes, 100µl 10mg/ml Herring Sperm DNA (Sigma) was precipitated with 10µl 3M Sodium Acetate pH5.2 and 500µl 100% ethanol. Tubes were incubated at -20°C for a minimum of 1hr, centrifuged at 18,000g for 30min at 4°C and washed with 80% ethanol. Pellets were resuspended in 120µl preheated hybridisation buffer (50% deionised formamide (Fluka)/ 2xSSC/ 5% dextran sulphate (Amersham)/ 10mM Tris pH7.4/ 0.1% Tween 20 (BDH)/ 0.2M Cysteamine (Sigma)) for sample tubes and 165µl for prehybridisation tubes (for WGTP microarrays, both tubes were resuspended in 165µl).

2.5.1.3 Hybridisation and washing

All sample and prehybridisation tubes were denatured at 70°C for 10min. 100µl from the prehybridisation tubes (140µl for the WGTP microarray) was immediately injected onto the microarray slide within Tecan chambers whilst the sample tubes were incubated at 37°C for 45min prior to injection of 100µl (140µl for the WGTP microarray). Samples were hybridised to the microarray at 37°C for 21hr (64hr for PCR product microarrays) with gentle agitation before being washed 15 times in 1xPBS/ 0.05% Tween20/ 2mM Cysteamine at 37°C for 1min, 5 times in 0.1xSSC at 54°C (52°C for PCR product microarrays) for 3min, 10 times in 1xPBS/ 0.05% Tween20/ 2mM Cysteamine at 25°C for 1min, before being rinsed in water at 25°C for 30sec prior to drying with nitrogen. Slides were scanned immediately after drying.

2.5.1.4 Scanning and image quantification

Fosmid and PCR product microarrays were scanned at a 10µm resolution and WGTP microarrays were scanned at a 5µm resolution on a G2565 scanner (Agilent Technologies). Fosmid and PCR product microarray images were quantified using GenePix Pro version 6.0 software (Axon Instruments) and the data analysed using a custom-made Excel spreadsheet. WGTP microarrays were quantified using BlueFuse for microarrays (BlueGnome).

2.5.2 Hybridisation and analysis of NimbleGen oligonucleotide microarrays (carried out by NimbleGen Systems)

Labelling, hybridisation and scanning were performed by NimbleGen Systems Inc. Briefly; 1µg genomic DNA (section 2.2.1) or Repli-G amplified flow sorted chromosomes (section 2.3.1.1) was denatured at 98°C in the presence of 1 O.D. of 50-Cy3- or 50-Cy5-labeled random nonamer (TriLink Biotechnologies) in 62.5mM Tris-HCl pH7.5/ 6.25mM $MgCl_2$/ 0.0875% ß-mercaptoethanol. The denatured sample was chilled on ice, then incubated with 100U (exo-) Klenow fragment (NEB) and 6mM each dNTPs for 2hr at 37°C. Reactions were terminated by addition of 0.5M EDTA (pH8.0) and the products precipitated with isopropanol and resuspended in water. The Cy3 and Cy5 labelled samples were combined (15µg each) and dried by vacuum centrifugation. The sample was rehydrated in 40µl of NimbleGen Hybridisation Buffer, denatured at 95°C for 5min, then cooled to 42°C. Hybridisations were carried out for 18hr at 42°C. The arrays were washed using a NimbleGen Wash Buffer System and immediately dried down by centrifugation. Arrays were scanned at 5µm resolution using the GenePix4000B scanner (Axon Instruments). Data were extracted from scanned images using NimbleScan 2.0 extraction software (NimbleGen Systems, Inc.)

2.6 Fluorescence in situ hybridisation

2.6.1 Preparation of slides for FISH hybridisation

2.6.1.1 Metaphase spreads

20ml of lymphoblastoid cell line culture was grown to confluency and subcultured. Approximately 24hr after subculturing, 100µl 3mg/ml Bromodeoxyuridine (BrdU;Sigma) in 2% ethanol was added and the cells incubated for 3hr at 37°C. 20µl 10mg/ml ethidium bromide and 40µl 10µg/ml colcemid (Invitrogen) was added prior to a further incubation at 37°C for 2hr. The cells were collected by centrifugation at 300g for 5min and the cells resuspended in 10ml 75mM KCl prewarmed to 37°C before incubation at 37°C for 12min. 3ml ice-cold fixative (3:1

methanol (BDH): glacial acetic acid (VWR International)) was added and the cells collected by centrifugation as before, then resuspended in 10ml fixative. The cells were collected and resuspended in fixative a further 2 times to remove cytoplasm and finally resuspended in approximately 5ml fixative. Metaphase suspensions were stored at -20°C in tubes sealed with parafilm for up to 3 years.

Two drops of metaphase suspension were dropped onto a glass slide and the slide incubated above a 50°C water bath to air dry. Slides were stored in a sealed box at room temperature for up to 7 days prior to use.

2.6.1.2 Extended chromatin fibres

Cells were collected from 2-3ml of an actively growing culture by centrifugation at 300g for 5min. The cells were washed twice in 3ml PBS, re-spun and finally resuspended in PBS to a concentration of 2-3x10^6 cells/ml.

Using a pipette, 10µl of cell suspension was spread over a 1cm area on the upper portion of a microscope slide and allowed to air dry. The slide was clamped to a Perspex block (made in-house) in a vertical position (Figure 2.3). To lyse the cells, 150µl lysis buffer (0.25M NaOH/ 27.4% ethanol) was allowed to pass along the slide by gravity flow until the meniscus stopped moving. A further 150µl 96% ethanol was then added and allowed to flow down the slide to remove the lysis buffer. The slide was gently separated from the Perspex block. This mechanical process allowed the DNA to be pulled into fibres and extend along the slide. The slides were air dried and stored in a box with desiccant for up to 6 months prior to use.

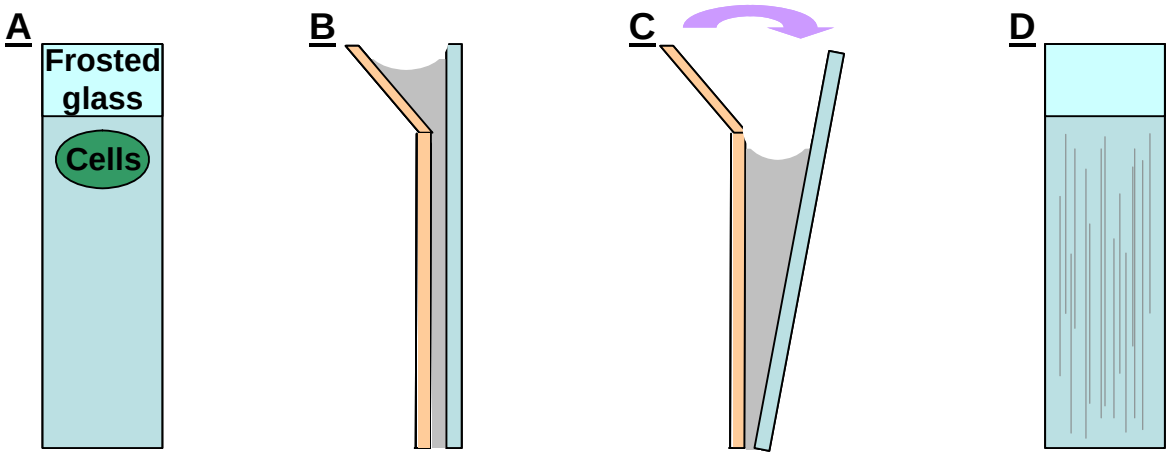


Figure 2.3 Schematic showing the production of extended chromatin fibres. A The cell spot was created on the upper portion of the microscope slide. B Cross section of the microscope slide (blue) clamped to the Perspex block (orange). The lysis buffer and ethanol (grey) pass along the microscope slide, extending the DNA from the cell spot into fibres. C Gentle separation of the slide from the holder allows the fibres to extend down the slide. D The fibres adhere along the length of the microscope slide.

2.6.2 Preparation of clone insert DNA

2.6.2.1 Preparation of fosmid clone insert DNA

Fosmid clone insert DNA was prepped from 40ml cultures using the Sigma PhasePrep BAC DNA Kit according to the manufacturer's instructions; cells were collected from an overnight 40ml culture by centrifugation at 4000g for 10min. 2ml Resuspension Solution was used to resuspend the cells prior to the addition of 2ml Lysis Solution. The tubes were inverted 5 times to allow mixing and incubated for 5min at room temperature. 2ml chilled Neutralisation Solution was added and the tubes inverted 8 times before incubation on ice for 5min. The tubes were centrifuged at 15,000g for 20 min at 4°C and the clear supernatant transferred to a fresh tube and this centrifugation step repeated. 3.6ml isopropanol was added to the clear supernatant, the tubes inverted and centrifuged at 15,000g for 20min at 4°C. The pellet was washed with 2ml 70%

ethanol and the pellet air dried briefly. The pellet was resuspended in 650µl Elution Solution followed by the addition of 1µl of 1/10 RNase cocktail and incubated at 60°C for 10min. 50µl 3M Sodium Acetate Buffer Solution pH7 and 120µl Endotoxin Removal Solution were added and the tubes incubated on ice for 5min, then warmed at 37°C for 5min prior to centrifugation at 16,000g for 3min. The clear upper phase containing the DNA was transferred to a clean tube and the lower blue phase containing endotoxins and other impurities was discarded. 700µl DNA Precipitation Solution and 2µl Pellet Paint (Novagen) were added and the DNA collected by centrifugation at 21,000g for 20min at 4°C. The pellets were washed with 500µl 70% ethanol and air dried prior to resuspension in 20µl water. DNA concentration and quality was assessed using an ND-1000 spectrophotometer (Nanodrop) and 1% agarose gel electrophoresis with ethidium bromide staining (Figure 2.4).

2.6.2.2 Preparation of plasmid clone insert DNA

5ml LB broth with 100µg/ml ampicillin was inoculated with bacterial glycerol stock and incubated at 37°C with shaking for 8hr. 1ml of this culture was used to inoculate 100ml of LB media with 100µg/ml ampicillin and incubated at 37°C with shaking overnight. Plasmid clone insert DNA was extracted from this culture using the Plasmid Maxi purification system (QIAGEN) according to the manufacturer's instructions. Briefly, bacterial cells were pelleted by centrifugation at 6000g for 15min at 4°C and resuspended in 10ml Buffer P1 (Resuspension Buffer; 50mM Tris-Cl pH8/ 10mM EDTA/ 100µg/ml RNase A). Following the addition of 10ml Buffer P2 (Lysis Buffer; 200mM NaOH/ 1% SDS) the cells were incubated at room temperature for 5min, then 10ml of ice-cold Buffer P3 (Neutralisation Buffer; 3M KOAc pH5.5) was added prior to incubation on ice for 20min. The samples were then centrifuged at 20,000g for 30min at 4°C and the supernatant removed and re-spun at 20,000g for 15min at 4°C. The supernatant was applied to a QIAGEN-tip 500 (already equilibrated with 10ml Buffer QBT (Equilibration Buffer; 750mM NaCl/ 50mM 3-[N-Morpholino] propanesulphonic

acid (MOPS) pH7/ 15% isopropanol/ 0.15% Triton X-100)) and washed with 2x30ml Buffer QC (Wash Buffer; 1M NaCl/ 50mM MOPS pH7/ 15% isopropanol). The DNA was eluted with 15ml Buffer QF (Elution Buffer; 1.25M NaCl/ 50mM Tris-Cl pH8.5/ 15% isopropanol), precipitated by the addition of 10.5ml isopropanol and collected by centrifugation at 15,000g for 30min at 4°C. The pellet was washed with 5ml 70% ethanol and centrifuged at 15,000g for 10min at 4°C. The pellet was air dried and dissolved in 500µl water. DNA concentration and quality was assessed using an ND-1000 spectrophotometer (Nanodrop) and 1% agarose gel electrophoresis with ethidium bromide staining (Figure 2.4).

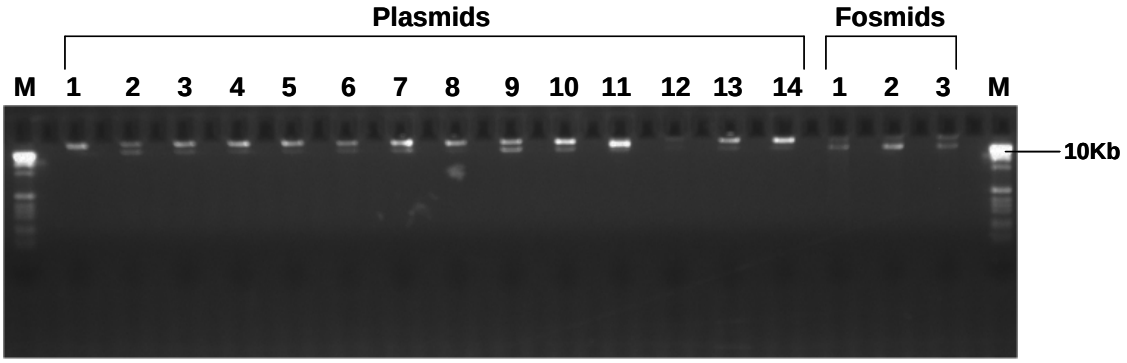


Figure 2.4 Isolated fosmid and plasmid insert DNA.

2.6.3 Fluorescence labelling of probes

2.6.3.1 Whole genome amplification labelling of clone DNA

10ng of prepped DNA (section 2.6.2) was amplified using the GenomePlex Complete Whole Genome Amplification Kit (Sigma) according to the manufacturer's instructions. Briefly; 10ng (1ng/µl) DNA was fragmented with 1µl 10xFragmentation Buffer at 95°C for 4min and immediately cooled on ice. The fragmented DNA was then used to create a library following the addition of 2µl 1xLibrary Preparation Buffer and 1µl Library Stabilisation Solution. The samples were heated at 95°C for 2min and cooled on ice prior to the addition of 1µl Library Preparation Enzyme. Samples were incubated at 16°C for 20min, 24°C

for 20min, 37°C for 20min and 75°C for 5min. For amplification of the library, 47.5µl nuclease-free water, 7.5µl 10xAmplification Mastermix and 5µl WGA DNA Polymerase were added and the samples subjected to temperature cycling of 95°C for 3min followed by 17 cycles of 94°C for 15sec and 65°C for 5min. Amplification was verified by running 6µl of the amplified library by gel electrophoresis on a 1.5% agarose gel with ethidium bromide staining.

1µl of the amplified DNA was labelled in a 25µl reaction with 2.5µl 10x Amplification Mix (special order from Sigma; A5604 without dNTPs), 1.5µl 1mM Biotin-16-dUTP (Roche), 1µl 1mM dTTP, 1.75µl WGA DNA polymerase (Sigma) and 17.25µl water. The samples were subjected to temperature cycling of 95°C for 3min followed by 17 cycles of 94°C for 15sec and 65°C for 5min. The fragment sizes of the labelled product were assessed by gel electrophoresis running 1µl on a 1% agarose gel with ethidium bromide (Figure 2.5).

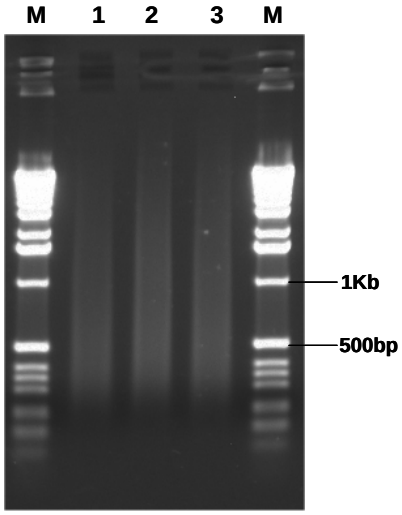


Figure 2.5 Amplified, labelled DNA for FISH.

To reduce the fragment sizes prior to hybridisation to the metaphase slides, 4µl 1µg/µl DNAse1 (Sigma) was added to the remaining 24µl of labelled DNA and incubated at 15°C for 1hr 30min. The fragment sizes were reassessed by gel electrophoresis as before. Optimal fragment sizes for FISH were considered to be 50-500bp in length (Figure 2.6).

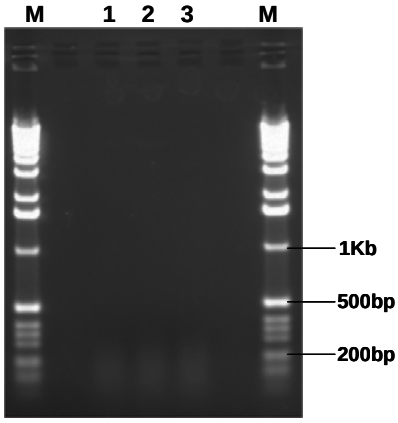


Figure 2.6 Amplified, labelled DNA after nick translation.

2.6.3.2 Nick translation of clone DNA

1µg of prepped plasmid or fosmid DNA (section 2.6.2) in 16µl water was added to 4µl biotin or dig nick translation mix (Roche) and incubated at 15°C for 1hr 30min. The reaction was stopped by the addition of 1µl 0.5M EDTA (pH8.0) and heating at 65°C for 10min. Fragment sizes were analysed by gel electrophoresis running 3µl of the reaction on a 1% agarose gel with ethidium bromide.

2.6.3.3 DOP PCR labelling of flow sorted chromosomes

2µl of a primary DOP reaction (section 2.3.1.2) was labelled in a 25µl secondary DOP reaction with 12µl DDW, 2.5µl 10xTAPS2 Buffer (250mM N-Tris (hydroxymethyl) methyl-3-amino-proane sulphonic acid pH9.3 (Sigma)/ 500mM KCl/ 20mM $MgCl_2$/ 10mM dithiothreitol/ 0.7% ß-mercaptoethanol/ 0.165% Bovine Serum Albumin), 2µl 20µM DOP primer (the same primer as used in the primary reaction), 2µl $^1/_2$T dNTPs (2.5mM each dATP, dCTP, dGTP and 1.25mM dTTP),

2.5µl 1mM Spectrum Green (Vysis), 1.25µl 1.25% W1 detergent (Sigma) and 0.25µl 5U/µl Amplitaq (PerkinElmer). Due to the higher annealing temperature used during a secondary DOP reaction the amplification was more specific than in a primary reaction. Samples were subjected to cycling conditions of 94°C for 4min followed by 35 cycles of 94°C for 1min, 62°C for 1min, 72°C for 1min 30sec followed by 72°C for 8min 30sec. Amplification was assessed by gel electrophoresis running 5µl on a 2.5% agarose gel with ethidium bromide.

2.6.4 Pre-treatment of metaphase spreads and DNA fibre slides

All slides were passed through an ethanol series (1min each in 70%, 70%, 90%, 90%, 100%) and allowed to air dry. Slides were pre-treated with 0.003% pepsin (Sigma)/ 0.01M HCl in 2xSSC at 37°C for 5min, followed by 3 washes in 2xSSC at room temperature for 5min, and a further ethanol series as before.

2.6.5 Hybridisation of FISH slides

For both metaphase spread and DNA fibre slides, 0.5µl of probe (section 2.6.3) was mixed with 1µl Cot1 DNA (Invitrogen) and 13.5µl hybridisation buffer (50% deionised formamide (Fluka)/ 2xSSC/ 10% dextran sulphate (Amersham)/ 10mM Tris pH7.4/ 0.1% Tween20 (BDH)). Samples were denatured at 65°C for 10min, prior to pre-annealing at 37°C for 50min before addition to the slides. Slides were denatured in 70% formamide/ 0.6xSSC at 65°C for 1min 30sec prior to quenching for 1min in ice-cold 70% ethanol and passing though an ethanol series as before. Hybridisations were performed under a coverslip sealed with rubber cement (Marubawerke GmbH & Co.) at 37°C overnight with humidity.

2.6.6 Detection of FISH slides

Indirectly labelled probes (sections 2.6.3.1 and 2.6.3.2) were detected using the schema detailed in Table 2.2. Slides were soaked in 2xSSC for 15min to remove the coverslip, washed in 2xSSC for 5min at 44°C, twice in 50% formamide/1xSSC for 5min at 44°C followed by a final wash in 2xSSC for 5min at

44°C. Slides were washed in 4xSSC/ 0.05% Tween20 (BDH) prior to antibody detection. Each slide was detected using 200µl of the relevant antibody (Table 2.2) in blocking buffer (1% blocking agent (Roche) in 4xSSC/ 0.05% Tween20/ 1µg/ml sodium azide) for 25min at 37°C and rinsed 3 times in 4xSSC/0.05% Tween for 4min at room temperature between layers.

	Biotin-16-dUTP		Digoxigenin-11-dUTP
1st layer	Avidin-Cy3 GE Healthcare 1µg/µl	Avidin-FITC Vector Laboratories 8µg/µl	Mouse-Anti-Digoxigenin Sigma; D8156 1/500 dilution of supplied stock
2nd layer	n/a	Anti-Avidin-FITC Vector Laboratories; SP2040 1/125 dilution of supplied stock	Goat-Anti-Mouse-Texas Red Invitrogen 8µg/µl

Table 2.2 Detection systems used for biotin and digoxigenin labelled probes.

Slides hybridised with directly labelled probes (section 2.6.3.3) were soaked in 2xSSC for 15min to remove the coverslip, then washed in 2xSSC for 5min at 42°C, twice in 50% formamide/0.5xSSC for 5min at 42°C followed by a final wash in 2xSSC for 5min at 42°C.

All slides were rinsed in 2xSSC at room temperature for 5min before staining for 3min in 0.2µg/ml DAPI (Sigma) in 2xSSC. Slides were then mounted with Citifluor mounting fluid (Citifluor, Ltd) and sealed with nail varnish.

2.6.7 Microscope analysis of FISH slides

Slides were visualised using an Axioskop microscope (Zeiss) with a CoolSNAP HQ camera (Photometrics) and narrow band pass filters (Chroma). Images were acquired at x100 magnification using SmartCapture X imaging software (Digital Scientific).

2.7 PCR

2.7.1 STS PCR

STS PCR was used to refine the translocation breakpoints to regions less than 5Kb prior to LR PCR amplification across the junction and also to create pools of PCR products for fosmid library screening. When mapping breakpoints, each STS primer pair was tested on whole genomic DNA from a normal individual as a positive control, water as a negative control, and both derivative chromosomes from the translocation and spanning BAC clone DNA as a control if available. When creating pools of products for library screening, each primer pair was individually used to amplify a product from genomic DNA (Promega) and tested on water as a negative control.

PCR primer pairs were designed to repeat-masked sequence using a perl script written in-house (Dimitris Kalaitzopoulos – Team 70). Oligonucleotides were designed to be 18-22bp in length with a melting temperature between 57°C and 63°C and GC content of 20-80%. Primer pairs were designed to amplify products of 90-120bp at the specified spacing along a target sequence.

15µl STS PCR reactions were performed in 96 well plates (Costar) on an MJ-Thermocycler. 5µl template DNA (50ng genomic DNA (Promega), 1/5 dilution of DOP amplified derivative chromosomes (section 2.3.1.2) or 5µl of colony DNA (created from a single colony resuspended in 50µl T0.1E) or water was amplified using 5.4µl 28% w/v Sucrose/cresol red, 2.5µl 10xNEB buffer (660mM Tris/ 167mM $(NH_4)_2SO_4$/ 67mM $MgCl_2$), 2.5µl 5mM each dNTPs, 0.495µl 0.5% Bovine Serum Albumin, 0.21µl 5% ß-mercaptoethanol and 0.18µl 5U/µl Amplitaq (PerkinElmer). 0.75µl of primer mix (100ng/µl of both forward and reverse primers) was added prior to temperature cycling of 94°C for 5min followed by 35 cycles of 94°C for 30sec, 58°C for 30sec, 72°C for 30sec followed by 72°C for 5min. 7.5µl of each reaction was run on a 2.5% agarose gel with ethidium

bromide. When creating PCR product pools for library screening, the whole 15µl reaction was run and the bands excised into 200µl T0.1E (10mM Tris, 0.1mM EDTA pH8).

2.7.2 Colony PCR

Colony PCR was used to verify that clones identified by library screening contained the translocation junction. The forward and reverse primers detailed in Table 2.3 were used to amplify products from clone DNA; Clones were picked and resuspended in 50µl T0.1E. 5µl of this DNA was used as template for the PCR reaction with 5µl (50ng) of patient DNA as a positive control and 5µl water as a negative control. Reactions were performed as detailed in Section 2.7.1 above.

Junction	Forward primer sequence	Reverse primer sequence
Derivative 7	GTAGTGATTCGGCCTTGCAT	TGGCCATATTTGGCTTTTTG
Derivative 13	TCCATTCATGTTGCTGCATT	GGAAGACAGGATGGATTCAAA

Table 2.3 Primers used to amplify products across the derivative chromosome 7 and 13 translocation junctions for patient t(7;13)(q31.1;q21.3).

2.7.3 LR PCR amplification of junction fragments

Oligonucleotide primers were designed using Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) to be 28-32bp in length, have a GC content of 40-60% and a melting temperature of 58-63°C.

25µl LR PCR reactions were performed in 96 well plates (Costar) on an MJ-Thermocycler. To 5µl of template (100ng genomic DNA (section 2.2.1)), 5µl Q solution, 2.5µl Hotstar buffer, 3.5µl 2mM each dNTPs, 0.5µl HotstarTaq DNA Polymerase, 0.5µl diluted Proofstart (0.8µl Proofstart, 1µl 10xProofstart buffer, 8.2µl water) and 7.6µl water were added. 0.2µl 100µM forward primer and 0.2µl 100µM reverse primer were added prior to temperature cycling of 95°C for 15min followed by 40 cycles of 95°C for 20sec, 57°C for 1min, 68°C for 10min followed

by 68°C for 10min. 5µl of each reaction was analysed using 0.6% agarose gel electrophoresis with ethidium bromide staining.

2.8 Sequencing

2.8.1 Sequencing of LR PCR products (Nik Matthews – Team 56)

Amplified junction fragment DNA (section 2.7.3) was prepared for sequencing using the ExoSAP method (Amersham). Cleaned fragments were sequenced from both ends using the di-deoxy chain termination method, with V.3.1 Big Dye Terminator chemistry (Applied Biosystems). Resulting sequencing reactions were analysed on 3700 ABI sequencing machines (Applied Biosystems).

2.8.2 End sequencing of fosmid clones (Teams 42 and 56)

Fosmid DNA was extracted using Millipore filter plates on a vacuum manifold and sequenced using V.3.1 Big Dye Terminator Cycle Sequencing Kit (Applied Biosystems) using a MJ thermocycler. Reactions were run on an ABI 3730 capillary sequencer (Applied Biosystems).

2.8.3 Full sequencing of fosmid clones (Teams 41, 53, 57, 116)

Random shotgun sequences of fosmid clones were generated from pUC19 plasmids with inserts of mainly 2–4 kb which were sequenced from both ends using the di-deoxy chain termination method. The resulting sequencing reactions were analysed on ABI 3730 sequencing machines (Applied Biosystems), and the data generated processed by a suite of in-house programs before assembly with the PHRED and PHRAP (http://www.phrap.org/) algorithms. The GAP4 program was used to assess and close sequence gaps.

2.9 Custom-made library production

2.9.1 In-house cosmid library

2.9.1.1 Extraction of high molecular weight DNA

$5x10^7$ cells were collected by centrifugation at 250g for 10min and resuspended in 5ml PBS. The cells were pelleted by centrifugation as before and resuspended in 1ml TE pH8 (10mM Tris/ 1mM EDTA). 10ml Lysis Buffer (10mM Tris pH8/ 0.1M EDTA/ 0.5% SDS/ 20µg/ml RNase (Sigma)) was added and the sample incubated at 37°C for 1hr. 55µl 20mg/ml Proteinase K (Sigma) was added and the sample incubated at 50°C for 3hr. The solution was cooled to room temperature and the DNA extracted with equal volumes of phenol/chloroform followed by chloroform. Each time, the two phases were separated by centrifugation at 5000g for 15min and the aqueous phase removed with a wide bore tip. DNA was precipitated with 0.2 volumes of 10M ammonium acetate and 2 volumes ethanol. DNA was spooled with a sterile loop, transferred to a tube containing 500µl 70% ethanol and spun at 5000g for 5min. The pellet was washed with 70% ethanol, air dried and resuspended in 10ml TE pH8.

2.9.1.2 Preparation of vector arms

Lawrist 16 cells were streaked onto LB agar containing 100µg/ml ampicillin and 20µg/ml kanamycin and incubated at 30°C until colonies were visible. A single colony was selected and the plasmid DNA collected (section 2.6.2.2). DNA was dissolved in 1ml TE (10mM Tris/ 1mM EDTA) and quantified using a TD-360 fluorometer (Turner Designs).

20µg of the DNA was digested in a 100µl reaction with 10µl 10xHSRE buffer (500mM Tris-HCl pH7.5/ 100mM $MgCl_2$/ 1500mM NaCl) and 9µl 10U/µl *Sca1* (NEB) at 37°C for 2hr. 1µl (200ng) of the digestion was checked on a 0.6% agarose gel alongside 200ng of the undigested DNA. A 5µl aliquot was removed as a *Sca1* control. To the remainder of the digestion, 276µl 1xLSRE buffer

(50mM Tris-HCl pH7.5/ 100mM $MgCl_2$) and 35µl 1U/µl Calf Intestinal Alkaline Phosphatase (CIAP; Roche) were added and the sample incubated at 37°C for 45min. 45µl 150mM Nitriloacetic Acid (NTA; Sigma) was added and the tube incubated for a further 25min at 68°C. Samples were then extracted with 450µl phenol/chloroform, then chloroform, each time re-extracting the organic phase with 450µl TE. Finally the aqueous phase was extracted with ether and the tube incubated at 68°C to evaporate any remaining ether. The DNA was precipitated with 45µl 3M sodium acetate and 1125µl ethanol at -20°C overnight. DNA was collected by centrifugation at 15,300g, the pellet washed with 70% ethanol, air dried and resuspended in 177.5µl TE. A 5µl aliquot was removed as a *Sca1*/CIAP control. 20µl 10xHSRE buffer and 7.5µl 20U/µl *BamH1* (NEB) were added and the tube incubated for 90min at 37°. DNA was extracted with equal volumes of phenol/chloroform, chloroform, and ether as described above. DNA was precipitated with $^{1}/_{10}^{th}$ volume sodium acetate and 2.5 volumes ethanol and resuspended in 100µl TE. A 5µl aliquot was removed as *Sca1*/CIAP/*BamH1* control.

45µl TE was added to all three 5µl control aliquots and the DNA extracted with 50µl phenol/chloroform and precipitated with 5µl sodium acetate and 125µl ethanol at -20° overnight. DNA was collected by centrifugation at 15,300g for 15min, washed with 70% ethanol, air dried and resuspended in 5µl TE and analysed using 0.6% gel electrophoresis with ethidium bromide staining.

2.9.1.3 Partial digestion of DNA

DNA (section 2.2.2) was divided into 7x150ng aliquots and partially digested using dilutions of *Mbo1* (NEB) and dam methylase (NEB). 150ng DNA in 18µl was added to 2µl 10xTAK buffer (300mM Tris-HCl pH7.9/ 600mM KOAc/ 90mM MgAc2/ 5mM DTT/ 3mg/ml Bovine Serum Albumin/ 800µM S-adenosyl methionine (SAM; Sigma)). 6µl was removed from each tube and retained as an

undigested control. 1µl of the appropriate enzyme mix from Table 2.4 was added to the relevant tube.

Tube name	Enzyme proportions	Enzyme mix
	Mbo1 **:dam methylase**	
BASIC	1:720	4µl *Mbo1* (0.1U/µl)
		36µl dam methylase (8U/µl)
1.3	1:2,320	4µl BASIC
		8µl dam methylase (8U/µl)
1.2	1:3,120	4µl BASIC
		12µl dam methylase (8U/µl)
1.1	1:7,920	1µl BASIC
		9µl dam methylase (8U/µl)
1.0	1:15,920	2µl 1.2
		8µl dam methylase (8U/µl)
dam	-	1µl dam methylase (8U/µl)

Table 2.4 Enzyme mixtures for partial digestion of genomic DNA.

All samples (including the undigested controls) were incubated for 2hr 30min at 37°C. 1µl 0.02U/µl CIAP (Roche) was added to tubes 1.0-1.3 and BASIC+ and the incubation continued for a further 30min at 37°C.

7µl was removed from each of the sample tubes and run with the undigested controls on a 0.3% agarose gel followed by ethidium bromide staining to assess for the best level of digestion (the fragment sizes averaging 36Kb).

11µl T0.1E (10mM Tris, 0.1mM EDTA) was added to tubes 1.0-1.3, BASIC+ and BASIC-. 2.2µl 150mM NTA was added to 1.0-1.3 and BASIC+ and all tubes incubated at 68°C for 20min prior to the addition of 1.2µl 5M NaCl and 60µl ethanol. Samples were precipitated overnight at -20°C.

2.9.1.4 Recovery of digested DNA

DNA for the selected digestion and BASIC+ and BASIC- tubes was collected by centrifugation at 15,300g for 15min, washed with 70% ethanol, air dried and

dissolved in 4µl T0.1E for the selected best digestion tube or 8µl for BASIC tubes.

2.9.1.5 Ligation of vector arms to digested DNA

2µl (400ng) vector arms were ligated to 50ng genomic DNA in an 8µl reaction with 0.8µl 10xligation buffer (400mM Tris-HCl pH7.9/ 100mM $MgCl_2$/ 10mM DTT), 0.7µl 400U/µl T_4 DNA Ligase (NEB) and 0.5µl 6mM ATP at 14°C overnight. To assess the background of non-recombinant clones, a control ligation was conducted where the genomic DNA was replaced with 4µl T1.0E. To assess the phosphatasing, the BASIC+ and BASIC- samples were split in half. The first half was ligated without arms and the second without arms and ligase. Both BASIC samples were run on a 0.3% agarose gel as before.

2.9.1.6 Packaging of ligated DNA

4µl of the ligation reaction (section 2.9.1.5) was added to a vial of Gigapack Gold III packaging extract (Stratagene) and incubated at room temperature for 2hr. The reaction was stopped by the addition of 500µl λ diluent (10mM Tris-HCl pH7.5/ 10mM $MgSO_4$) and 132µl 5xSM (500mM NaCl/ 50mM $MgSO_4$/ 250mM Tris-HCl pH7.5/ 0.05% gelatine/ 50% glycerol).

2.9.1.7 Preparation of plating cells

40ml LB broth was inoculated with a single colony of DH5αMCR E.coli and incubated at 37°C with shaking overnight. Cells were pelleted at 3,200g for 15min and resuspended in 20ml 10mM $MgSO_4$.

2.9.1.8 Plating of library

Varying strengths of packaged phage and λ diluent (ranging from 1:5 to 1:100 dilutions) totalling 100µl were added to 100µl plating cells and incubated at room temperature for 20min. The infected cells were then diluted with 1ml LB broth and incubated at 37°C for 45min to allow expression of antibiotic resistance. The cells were then pelleted by centrifugation at 3,300g for 2min. The majority of the

supernatant was removed, leaving approximately 30µl remaining for resuspension of the cells which were then plated on LB agar with 30µg/ml kanamycin and incubated at 37°C overnight.

2.9.2 EpiFOS fosmid library

Flow sorted derivative chromosomes were used to generate custom-made fosmid libraries using the EpiFOS™ Fosmid Library Production Kit (Epicentre) according to the manufacturer's instructions.

2.9.2.1 Size modification of DNA

Flow sorted chromosomes (section 2.2.2) were resuspended in 200µl TE (10mM Tris-HCl (pH7.5)/ 1mM EDTA). 100ng of DNA was run on a 0.3% agarose gel using fosmid control DNA to assess size. To reduce its size, the DNA was sheared though a 21 gauge syringe needle (Microlance), and a further 100ng of DNA checked gel electrophoresis as before. The DNA was then concentrated to less than 52µl using an YM-30 microcon column (Millipore).

2.9.2.2 End-Repair of sheared DNA

In an 80µl reaction, the sheared DNA was mixed with 8µl End-Repair 10xBuffer, 8µl 2.5mM each dNTP mix, 8µl 10mM ATP and 4µl End-Repair Enzyme Mix and incubated at room temperature for 45min. Following the addition of blue loading buffer, the samples were incubated at 70°C for 10min to inactivate the enzyme.

2.9.2.3 Size selection of end-repaired DNA

End-repaired DNA was separated using a 0.3% low melting point agarose (Sigma) gel. A gel slice containing DNA of approximately 30-40Kb in size was excised from the gel and melted at 70°C for 15min. Following the addition of pre-warmed GELase 50x Buffer and GELase Enzyme Preparation (to a final concentration of 1x buffer and 1U enzyme per 0.1g excised agarose) the samples were incubated at 45°C for 3hr, followed by 70°C for 10min. 500µl aliquots were chilled on ice for 5min and then centrifuged at 9,300g for 20min.

The upper 90-95% of supernatant was transferred to a clean tube and precipitated with 1/10th volume of 3M Sodium Acetate pH7 and 2.5 volumes 100% ethanol at room temperature for 10min. The DNA was collected by centrifugation at 16,000g for 20min, washed twice with 70% ethanol, air dried and resuspended in 200µl TE buffer. The DNA was concentrated using an YM-30 microcon column (Millipore).

2.9.2.4 Ligation of size-selected DNA

10µl ligation reactions were performed using the size-selected DNA, 1µl 10xFast-Link Ligation Buffer, 1µl 10mM ATP, 1µl 0.5µg/µl pEpiFOS-5 Vector and 1µl Fast-Link DNA Ligase and incubation at room temperature for 2hr followed by 70°C for 10min.

2.9.2.5 Preparation of E.coli host strain

2 days prior to the packaging reactions EPI100^{TM}-T1^{R} cells were streaked onto LB agar plates and the plate incubated at 37°C overnight. The day preceding the packaging reaction, a single colony was picked from this plate and used to inoculate 50ml LB broth supplemented with 10mM $MgSO_4$ and incubated at 37°C overnight with gentle shaking. The day of the packaging reactions, 5ml from the overnight culture was used to inoculate 50ml LB broth supplemented with 10mM $MgSO_4$ and incubated at 37°C with gentle shaking to an OD_{600} of 0.8-1.0.

2.9.2.6 Packaging of ligated DNA

1 tube of MaxPlax Lambda Packaging Extract was thawed on ice and 25µl of the extract was transferred to the tube containing the 10µl ligation reaction; the other 25µl was returned to -70°C. After incubation of the sample tube at 30°C for 90min, the remaining 25µl of packaging extract was added, and the tube incubated for a further 90min at 30°C. After the second incubation, 940µl of Phage Dilution Buffer (10mM Tris-HCl pH8.3/ 100mM NaCl/ 10mM $MgCl_2$) and 25µl chloroform were added and the packaged phage stored at 4°C.

2.9.2.7 Plating of library

Before plating the whole library, a series of plating dilutions were performed to determine the optimum plating strength. In all cases, the packaged phage was diluted with Phage Dilution Buffer. 10µl of the diluted packaged phage was added to 100µl of the prepared EPI100^{TM}-T1^{R} cells (section 2.9.2.5) and incubated at 37°C for 20min prior to plating on LB agar plates with 12.5µg/ml chloramphenicol and incubation at 37°C overnight.

2.10 Generation of fosmid library filters

Polygrid filters were printed for all libraries created. Round filters were only created for the patient t(7;13)(q31.1;q21.3) library. For all libraries created, clones were picked from round agar plates into LB broth with 12.5µg/ml chloramphenicol in 384 well plates (Genetix) by a robot developed in-house and incubated at 37°C overnight with shaking.

2.10.1 Printing of polygrid filters from 384 well plates (Mark Maddison – Team 63)

Clones were gridded from 384 well plates onto Hybond-XL filters (Amersham) on top of LB agar plates with 12.5µg/ml chloramphenicol using a robot developed in-house with Black Magic Ink (Higgins) used to highlight the spot locations. Filters were incubated overnight at 37°C.

2.10.2 Replication of clones onto round filters from agar plates

Round agar plates with the plated library clones were used to generate lift filters for library screening. Hybond N+ filters (Amersham) were placed on top of the library plates and then moved to fresh agar plates with 12.5µg/ml chloramphenicol and incubated at 37°C overnight.

2.10.3 Preparation of clone DNA on filters for hybridisation

For both polygrid and round filter systems, the filters were placed colony side up on 3MM paper (Whatman) saturated with 10% SDS for 5min, then transferred to

paper saturated with denaturing solution (0.5M NaOH/ 1.5M NaCl) for 10min before being transferred to fresh paper and allowed to air dry for 15min. Filters were then subjected to a series of washes with gentle shaking; 10xNeutralising Solution (0.5M Tris-Cl pH7.7/ 1.5M NaCl) for 5min twice, 1xNeutralising Solution for 5min, 2xSSC/ 0.1% SDS for 5 min, 2xSSC for 5min and 100mM Tris-Cl pH7.4 for 5min before being allowed to air dry. DNA on the filters was crosslinked for 2min on a 90W UV-B transilluminator TL-312A (Spectroline) prior to use in hybridisations.

2.11 Radioactive screening of fosmid library filters

2.11.1 Generation of probe for library screening

2.11.1.1 Generation of probe by Inter-Alu PCR

100ng of spanning BAC DNA was amplified in a 25µl reaction using 2.5µl 10xNEB Buffer (670mM Tris-HCl pH8.8/ 166mM $(NH_4)_2SO_4$/ 67mM $MgCl_2$), 2.5µl 10xdNTP-C (5mM each dATP, dGTP, dTTP), 0.6µl (3000Ci/mmol) ^{32}P dCTP (Amersham), 0.825µl 5mg/ml Bovine Serum Albumin, 2.5µl 100ng/µl ALE1 primer (GCCTCCCAAAGTGCTGGGATTACAG), 2.5µl 80ng/µl ALE3 primer (CCAYTGCACTCCAGCCTGGG) 0.35µl 5% ß-mercaptoethanol and 0.3µl 5U/µl Amplitaq (PerkinElmer). Samples were subjected to temperature cycling of 94ºC for 5min followed by 30 cycles of 93ºC for 1min, 65ºC for 1min, 72ºC for 5min followed by 72ºC for 5min under mineral oil (Sigma).

125µl 20xSSC, 125µl Cot1 DNA (Invitrogen), 25µl radiolabelled probe and 225µl water were mixed together and incubated at 100ºC for 5min, prior to incubation on ice for 2min. The completed probe was then added to the pre-hybridised filters.

2.11.1.2 Generation of probe by STS PCR

A secondary round of PCR was performed to label products obtained in section 2.7.1. 1µl of the primary product was labelled in a 15µl reaction with 0.6µl 100ng/µl both forward and reverse primer mix, 1.5µl 10xNEB Buffer (670mM Tris-HCl pH8.8/ 166mM $(NH_4)_2SO_4$/ 67mM $MgCl_2$), 9.9µl T0.1E (10mM Tris, 0.1mM EDTA), 0.6µl dNTP-C (5mM each dATP, dGTP, dTTP), 0.5µl (3000Ci/mmol) ^{32}P dCTP (Amersham), 0.495µl 5mg/ml Bovine Serum Albumin, 0.21µl 5% ß-mercaptoethanol, 0.18µl 5U/µl Amplitaq (PerkinElmer). Samples were subjected to temperature cycling of 94ºC for 5 min followed by 25 cycles of 94ºC for 30sec, 55ºC for 30sec, 72ºC for 30sec followed by 72ºC for 5min under mineral oil (Sigma).

Samples were denatured at 94ºC for 5min prior to quenching on ice for 2min. The probes were then pooled on ice prior to addition to the pre-hybridised filters.

2.11.2 Hybridisation of radioactive DNA probe to fosmid library filters

Filters were pre-hybridised in filtered hybridisation buffer (4mg/ml Ficcoll (Sigma)/ 4mg/ml Bovine Serum Albumin (Sigma)/ 4mg/ml polyvinylpyrrolidone (Sigma)/ 6XSSC/ 50mM Tris pH7.4/ 200mg/ml Dextran Sulphate (Amersham)/ 1% sarkosyl) at 65°C at least 2hr prior to the addition of the radioactive probe. Following the addition of the probe, the filters were incubated at 65°C overnight with gentle shaking.

2.11.3 Washing and detection of fosmid library filters

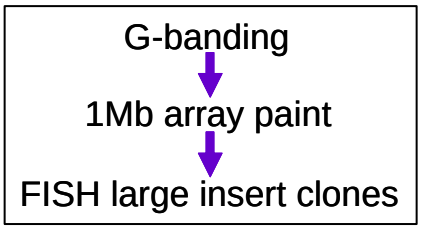
Hybridised filters were rinsed twice in 2xSSC for 5min with gentle shaking, washed in 0.5xSSC/ 1% sarkosyl washing solution at 65°C twice for 30min with gentle shaking and finally rinsed for 5min in 0.2xSSC with gentle shaking prior to laying out and wrapping in clingfilm. Filters were exposed to X-ray film (Fuji) overnight and developed using a Kodak processor.

3 Results: Investigation of chromosome rearrangements by array painting, comparative genomic hybridisation and PCR analysis

3.1 Introduction

Approximately 0.087% of all live births have a balanced reciprocal translocation (Jacobs 1981) leading to a 6.1% risk of a serious congenital abnormality (Warburton 1991). Routine Geimsa banding analysis of a patient's karyotype can identify the chromosomes involved in the rearrangement if the chromosomes involved are altered in size or band content. After successful identification of the chromosome translocation by karyotyping, the breakpoints have conventionally been mapped either by fluorescence *in situ* hybridisation using genomic insert clones or by isolating the derivative chromosomes in somatic cell hybrids and extracting the genomic DNA for restriction enzyme digestion. The digested DNA can be separated on agarose gels and blotted onto nitrocellulose membranes so that DNA fragments that span the breakpoints can be identified by Southern blotting. These fragments can then be cloned and sequenced to obtain the breakpoint junction sequence. This approach is lengthy and of the many different translocation breakpoints present in the population, fewer than fifty* have been sequenced and published proving the need for a more rapid technique for breakpoint isolation. This chapter demonstrates how DNA microarrays and PCR can be used to more efficiently map and amplify chromosome rearrangement breakpoints rapidly for sequencing.

*(Bodrug et al. 1987; Bodrug et al. 1991; Giacalone and Francke 1992; Arai et al. 1994; Budarf et al. 1995; van Bakel et al. 1995; Toriello et al. 1996; Holmes et al. 1997; Ishikawa_Brush et al. 1997; Kehrer-Sawatzki et al. 1997; Krebs et al. 1997; Kurahashi et al. 1998; Roberts et al. 1998; Yoshiura et al. 1998; Ikegawa et al. 1999; Holder et al. 2000; Matsumoto et al. 2000; Millar et al. 2000; Bonaglia et al. 2001; Nothwang et al. 2001; Willett-Brozick et al. 2001; Duba et al. 2002; McMullan et al. 2002; Spitz et al. 2002; Sugawara et al. 2002; Vervoort et al. 2002; David et al. 2003; Hill et al. 2003; Jeffries et al. 2003; Nimmakayalu et al. 2003; Shoichet et al. 2003; Gotter et al. 2004; Rodriguez-Perales et al. 2004; Bocciardi et al. 2005; Borg et al. 2005; Klar et al. 2005; Schule et al. 2005; Velagaleti et al. 2005; Haider et al. 2006; Mansouri et al. 2006; Tagariello et al. 2006)



Flow chart describing the progress of breakpoint mapping throughout each stage of this chapter.

The three chromosome rearrangement patients described in this chapter have been studied previously to identify translocation spanning clones (Gribble et al. 2005). Briefly, initial identification of constitutional apparently balanced reciprocal translocations by Geimsa banding analysis at Wessex Regional Genetics Laboratories was followed by array painting and array CGH using a 1Mb resolution microarray. Array painting resolved each translocation breakpoint to an interval of approximately 1Mb, and clones in these regions were selected from the human genome project 'Golden Path' clone set for FISH analysis. Hybridisation of clones across these regions to patient metaphase spreads revealed a change in the localisation of the FISH probe from one derivative chromosome to the other; sequential clones hybridised to one derivative chromosome, showed a split signal on both derivative chromosomes or hybridised to the other derivative chromosome. The translocation breakpoint was contained within the clone showing a hybridisation signal on both derivative chromosomes. Details of the karyotypes, translocation spanning BAC clones and any additional complexity identified can be seen in Table 3.1.

Translocation karyotype	Spanning clones	Accessioned position (bp)	Additional complexity
46,XY,t(2;7)(q37.3;p15.1)	RP11-680O16	2; 236,390,125 to 236,562,868	dup(3)(p26.3;p26.3)pat
	CTA-471E18	7; 30,853,334 to 31,010,213	
46,XX,t(3;11)(q21;q12)	RP11-529F4	3; 130,599,785 to 130,768,338	none found
	RP11-855O10	11; 61,100,132 to 61,285,897	
46,XY,t(7;13)(q31.3;q21.3)	RP11-384A20	7; 121,121,492 to 121,279,437	inv(11)(p15.3;p15.5)pat
	RP11-360I23	13; 70,947,118 to 71,124,154	

Table 3.1 Previously published translocation spanning clones for the three patients whose breakpoints were previously investigated using array based methodologies (Gribble et al. 2005). Clone positions were taken from Ensembl Build 35.

In order to refine the breakpoint containing regions further and sequence across the translocation breakpoints of these patients a number of different array based methodologies have been investigated and the findings reported below. In addition, further complexity within the patients' chromosomes has been investigated in this chapter using array CGH as previous studies have shown that 'balanced' rearrangements are often accompanied by cryptic imbalances either at the breakpoints or elsewhere in the genome (Kumar et al. 1998; Astbury et al. 2004; Patsalis et al. 2004).

3.2 Summary of methods

During the final stages of the sequencing of the human genome a fosmid library was used to assess the quality of the reference sequence. The library was created using DNA from an individual not used in generating the reference sequence. As part of the quality control process, both ends of 750,000 fosmid inserts were sequenced and the data generated used to map each fosmid relative to the reference sequence (IHGSC 2004). We used this fosmid resource to construct increased resolution microarrays to investigate chromosome rearrangement breakpoints. Whilst this fosmid resource is highly redundant (750,000 fosmids represent, on average, a 10 times coverage of the human genome), the fosmid map is not contiguous along every chromosome and gaps are still present. Array painting using flow sorted derivative chromosomes from the patient onto these fosmid microarrays mapped the breakpoints at an increased resolution compared with the spanning BAC clones previously identified. Array CGH using DNA from the patient against a reference DNA onto these fosmid microarrays probed the surrounding area for any cryptic imbalance.

Once the breakpoint area had been refined by array painting on the custom-made fosmid microarray, STS PCR was used to further narrow down the breakpoint region. Primer pairs were designed at intervals to unique regions across the genomic region of interest and derivative chromosomes were used as template DNA. A switch from a positive result on one derivative chromosome to a positive result on the other derivative delineates the breakpoint. In practice, two

rounds of PCR were sufficient to refine the breakpoint sufficiently for the translocation junction to be amplified by Long Range PCR. The first round of STS PCR was performed using primer pairs spaced approximately every 10-20Kb, and once a breakpoint spanning region had been identified, a further round of primer pairs could be designed every 1-3Kb within it, so in principle refining the breakpoint to within a 1-3Kb region.

To amplify a product containing the translocation junction sequence by LR PCR, primers were designed at intervals across the refined regions. The forward primers from one chromosome were paired with the reverse primers from the other chromosome and genomic DNA from the patient used as template (Figure 3.1).

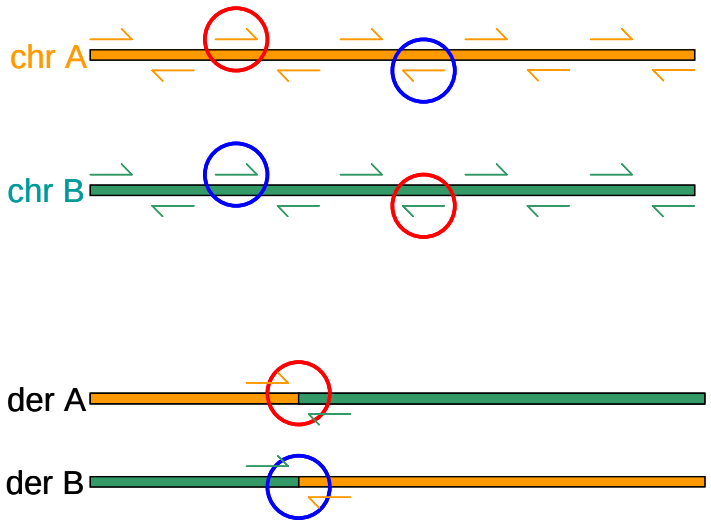
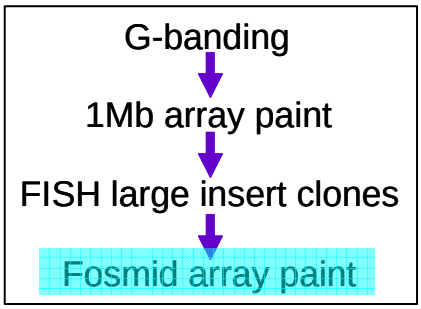


Figure 3.1 Diagrammatic representation of the principle behind long range PCR. Pairing a forward primer designed to one chromosome with the reverse primer designed to the other chromosome will amplify across the derivative chromosome junction.

After successful amplification of the breakpoint containing region by LR PCR the PCR product was sequenced. The resulting sequence was compared with the human reference sequence available in NCBI Build 35 to identify the precise breakpoints. The surrounding sequence was then studied in further detail as discussed in Chapter 4.

3.3 Results; Refinement of chromosome translocation breakpoints using custom-made fosmid microarrays



A high resolution custom-made fosmid array was constructed to study the translocation breakpoints of the three patients' rearrangements. The array consisted of 2244 elements; 327 large insert clones (PACs and BACs) selected at 10Mb intervals across the human genome spotted in duplicate, 517 fosmids spotted in triplicate to cover genomic regions of interest for the rearrangements being investigated and 5 Drosophila clones spotted in triplicate as controls to allow for quantitation of background hybridisation. To give the best possible coverage for the area of interest, fosmid clones were selected at full redundancy from the 'Fosmid End Pairs' track on the UCSC website (http://genome.ucsc.edu/) according to NCBI build 35 of the human genome. The clones were DOP PCR amplified then re-amplified by an amino-linking PCR to attach an amino group to the 5' end of each product to allow for covalent bonding to the microarray slide. The fosmid microarrays were used in array painting and CGH hybridisation experiments as described in the methods section.

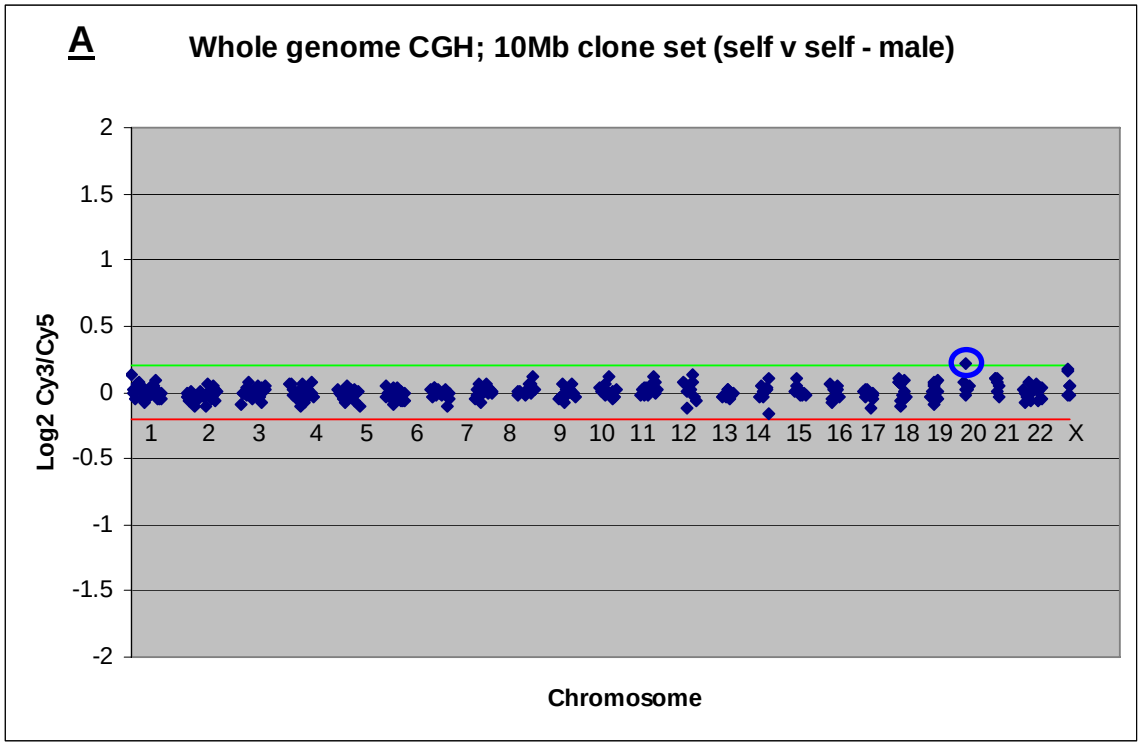
All fosmid clones were spotted in triplicate and the $\log_2$ ratio (Cy3/Cy5 intensities) of each spot within a trio was compared with the other two spots. If all 3 spots were within 2 standard deviations of the mean, the ratio was used for analysis. If the standard deviation of all 3 clones was found to be greater than 2, the standard deviation for all 3 combinations of 2 spots was calculated and if 2 clones within each trio were found to be within 2 standard deviations of the mean, they were used for the analysis, and the other spot discounted. Any clones where any 2 of the 3 spots fell outside 2 standard deviations of the mean were excluded from further analysis.

3.3.1 Testing the custom-made fosmid microarray

Prior to using the fosmid microarray for breakpoint mapping of patient rearrangements, the clones on the array were tested using a self versus self hybridisation and a test sample with known amplifications and deletions. Using a $\log_2$ scale, a clone showing a heterozygous deletion will report a ratio close to -1 and a clone showing a duplication will report a ratio close to 0.58. Clones which reported unexpected ratios in these experiments were excluded from further analyses. Thresholds were set by adding or subtracting 4 standard deviations of the autosomal clone $\log_2$ ratios to the mean. Clones which displayed ratios above or below these thresholds were deemed to be amplified or deleted respectively.

3.3.1.1 Self versus self hybridisation

The self versus self experiment was performed using a pool of 20 normal male DNAs. One BAC clone on chromosome 20 (RP1-64G16) was identified as showing a significantly high ratio. All fosmid clones (Figure 3.2) were seen to lie close to a $\log_2$ ratio of zero indicating that there was no difference in hybridisation between samples with the exception of G248P87165E3 on chromosome 5 and G248P81609G9 on chromosome 11. Because the samples hybridised to the microarray were the same, the discrepancy in hybridisation to these clones must have arisen from a dye bias between the Cy3 and Cy5 dCTP labelled samples. RP1-64G16, G248P87165E3 and G248P81609G9 were excluded from further analyses on the basis of these results.


A
Whole genome CGH; 10Mb clone set (self v self - male)
Log2 Cy3/Cy5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
Chromosome

B

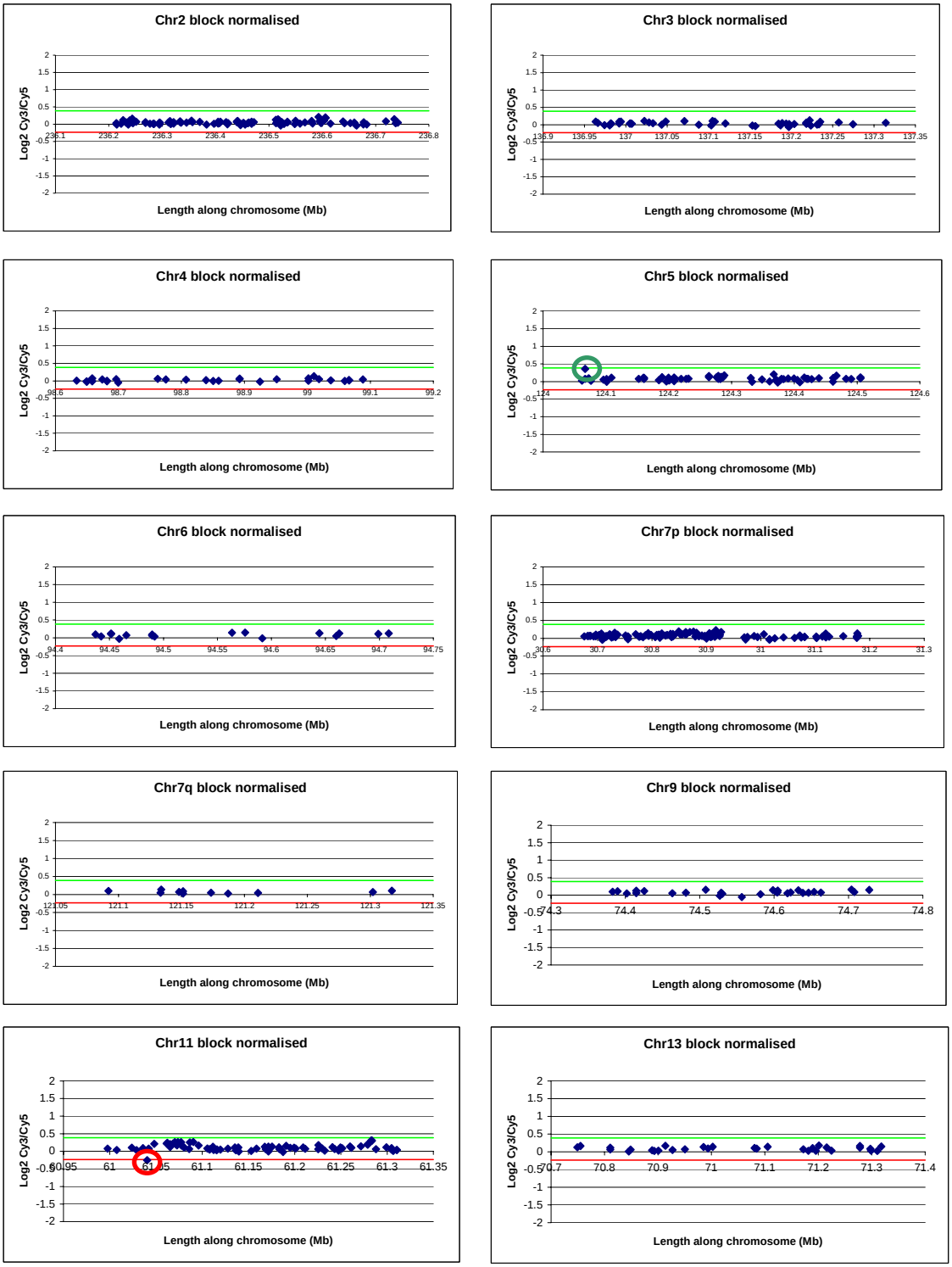


Figure 3.2 Self versus self CGH hybridisation onto custom-made fosmid microarray; A 10Mb clones, B fosmid clone results. RP1-64G14 G248P87165E3 G248P81609G9.

3.3.1.2 769p hybridisation

To investigate whether the clones on the microarray in general would respond correctly to a change in copy number, a primary renal cell adenocarcinoma cell line, 769p, well characterised with respect to copy number gains and losses was hybridised to the microarray.

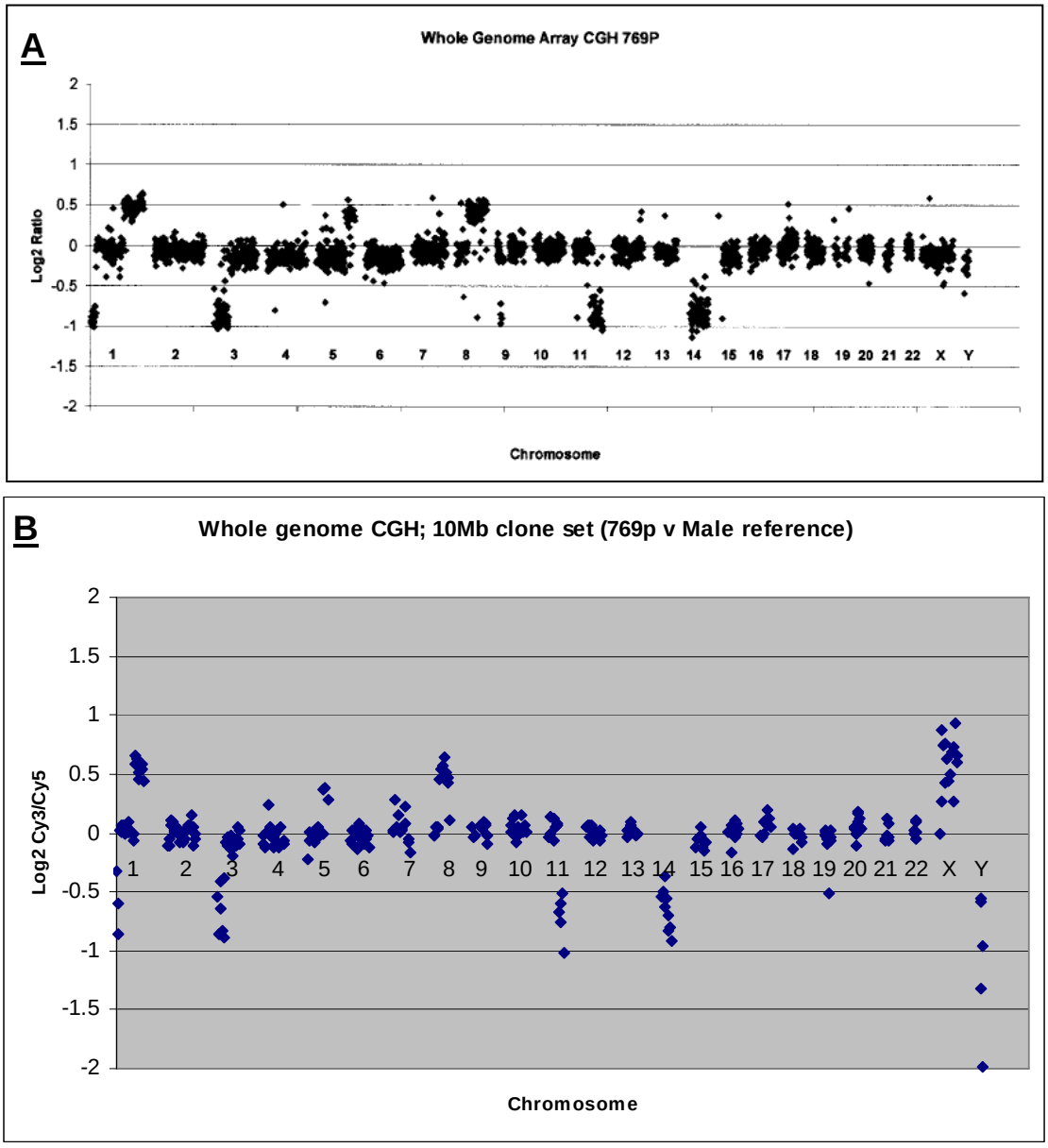


Figure 3.3 ***A Taken from Fiegler et al 2003 (Fiegler et al. 2003a). Array CGH profile of renal cell carcinoma cell line 769p against a female reference on a 1 Mb resolution whole genome microarray. B Results for large insert clones at 10Mb resolution in a 769p versus male reference hybridisation on custom-made fosmid microarray.***

The large insert clones at 10Mb resolution on the custom-made fosmid microarray reported regions of copy number gain (1q, 5q, 8q) and copy number loss (1p, 3p, 9p, 11q, 14) comparable to those reported by Fiegler *et al* on a 1Mb resolution microarray (Fiegler et al. 2003a) (Figure 3.3).

Whilst none of the regions covered by the fosmids on the array were in areas of amplification or deletion within the cell line (Figure 3.4) this 769p versus reference hybridisation was used to verify that none of the fosmids responded to regions of gain or loss within the cell line. There were 3 individual fosmid clones which reported a significant shift in ratio; G248P8541B6 which reported a gain on chromosome 7p and clones G248P87179H5 and G248P88359H5 which reported losses on chromosome 11. As these fosmid clones lay in regions of high redundancy of fosmid clones and as all other overlapping clones behaved as expected it was likely that these clones were mispicked or mismapped.

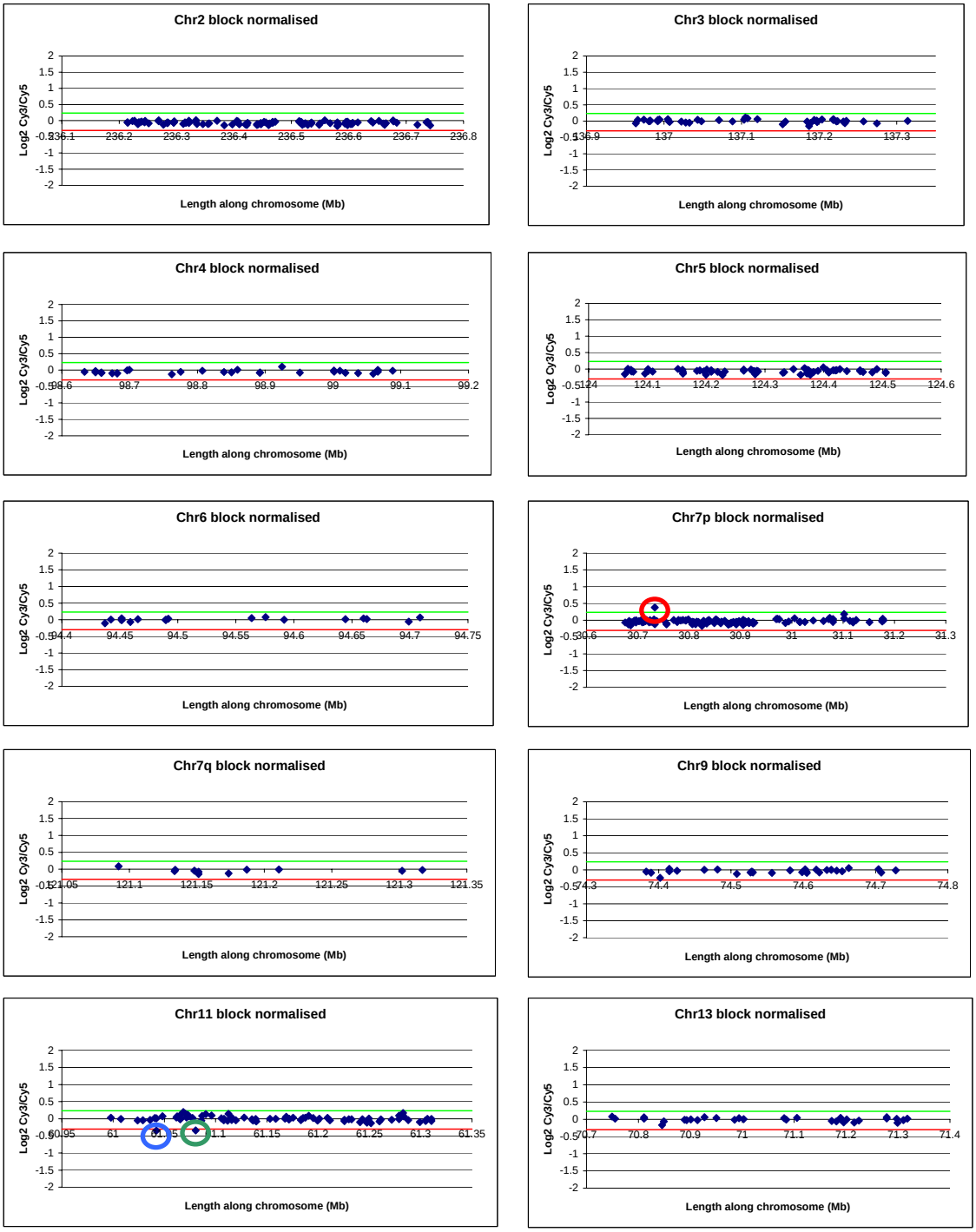
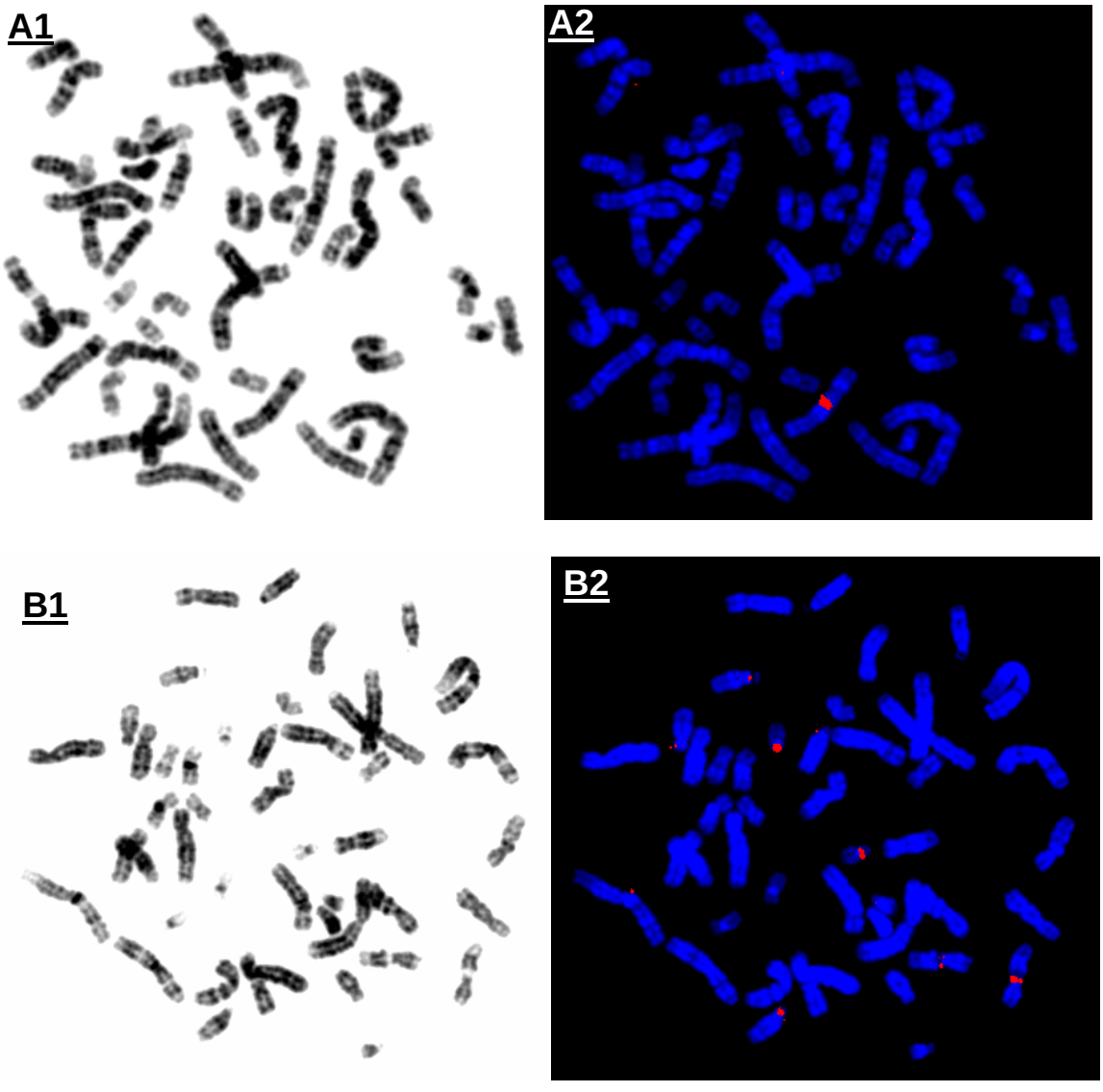


Figure 3.4 CGH results for renal cell carcinoma cell line on custom-made fosmid microarray for fosmid clones. G248P8541B6 G248P87179H5 G248P88359H5.

The three clones showing unexpected ratios were mapped by FISH to verify their chromosomal locations (Figure 3.5). G248P8541B6 hybridised to Xq13.1-Xq13.3

not 7p as expected. The elevated ratio seen by this clone after the CGH experiment was also seen for the other chromosome X clones spotted on the microarray. G248P87179H5 gave signals on the acrocentric chromosome centromeres (13, 14, 15, 21, 22). G248P88359H5 showed a signal on 11q13.1-11q13.4 as well as the centromere of chromosome 3 and the acrocentric chromosomes. The FISH data unequivocally explained the elevated ratio results seen after microarray hybridisation, and these clones were excluded from further analysis.



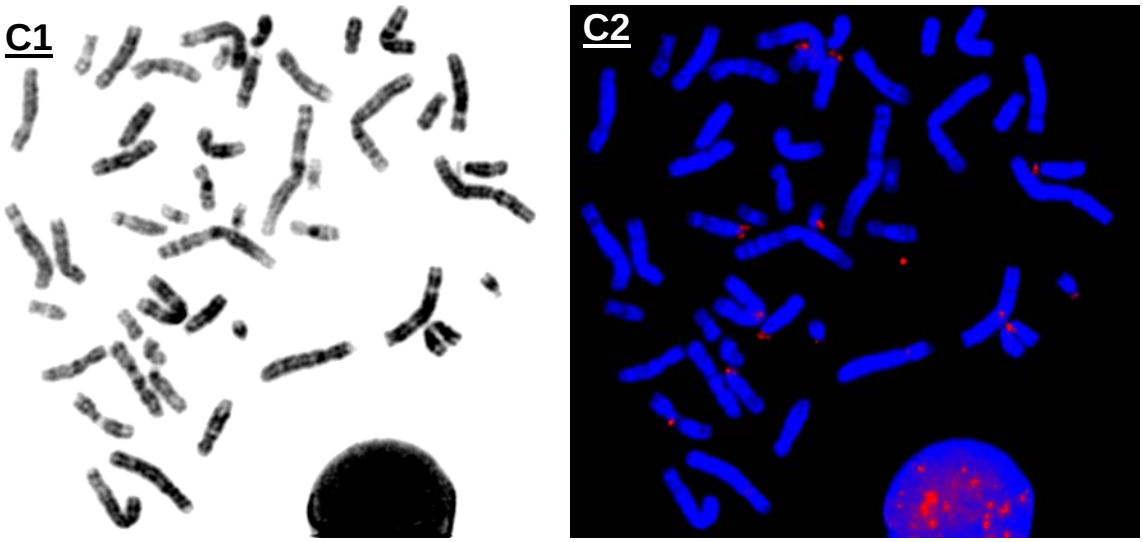


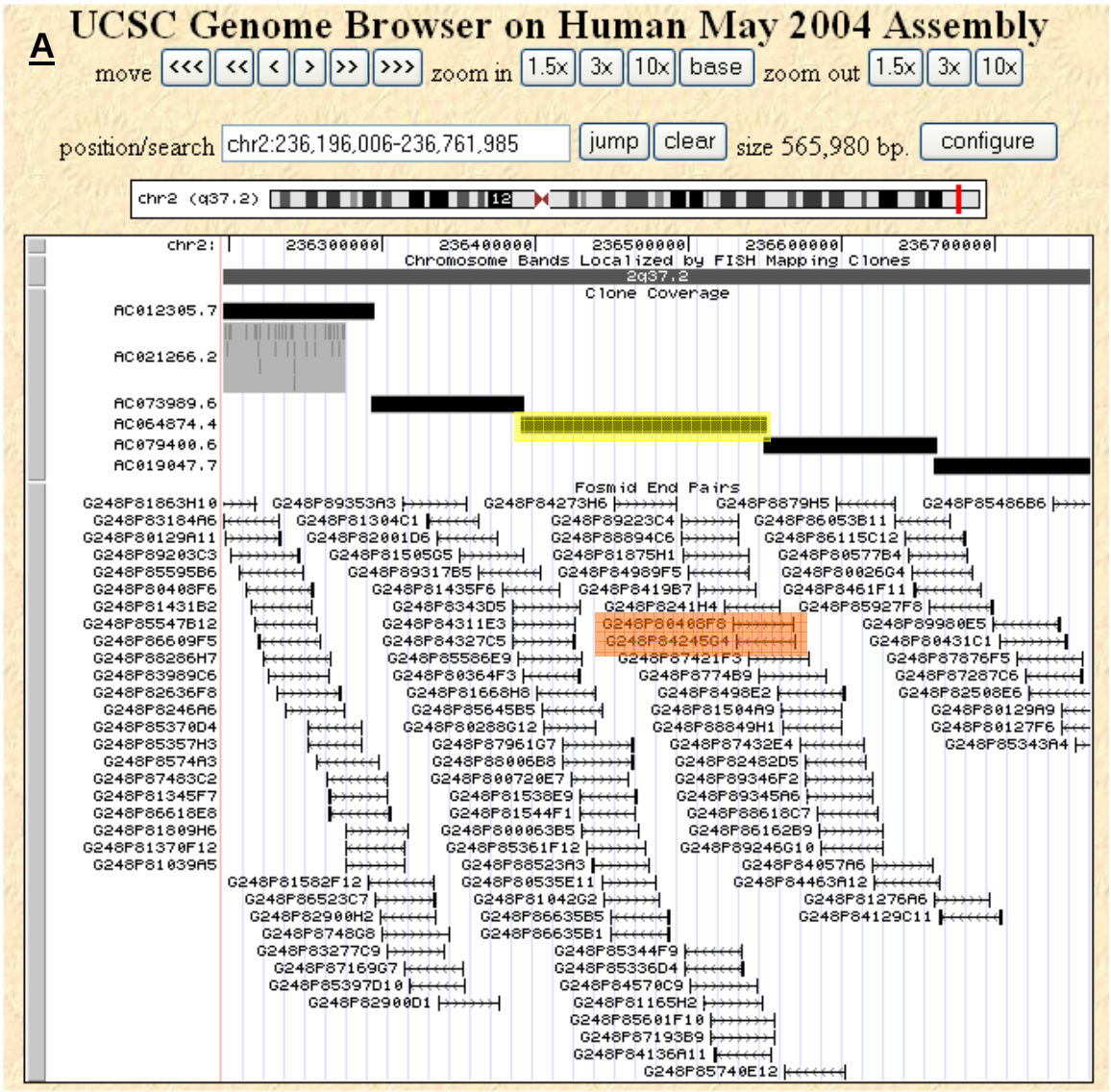
Figure 3.5 FISH results for clones showing unexpected ratios in a 769p versus reference array CGH hybridisation. 1 inverted DAPI banding of metaphase chromosomes, 2 metaphase FISH image. Results for A G248P8541B6 on Xq13.1-Xq13.3, B G248P87179H5 on the centromeres of acrocentric chromosomes and C G248P88359H5 on 11q13.1-11q13.4 and the centromeres of chromosomes 3, 13, 14, 15, 21 and 22.

3.3.2 Array painting analysis of patient derivative chromosomes on custom-made fosmid microarrays

Previous FISH analysis identified spanning BAC clones for all the three patient translocation breakpoints (Gribble et al. 2005). To further refine the approximately 150Kb regions defined by these spanning clones, derivative chromosomes were array painted onto the custom-made DNA fosmid microarray. This method was used to refine both breakpoints for patients t(2;7)(q37.3;p15.1) and t(7;13)(q31.3;q21.3), but only the chromosome 11 breakpoint for patient t(3;11)(q21;q12) as the chromosome 3 breakpoint had been refined to 2.2Kb prior to the start of this project.

3.3.2.1 Patient t(2;7)(q37.3;p15.1)

Chromosome 2 and 7 breakpoint spanning regions were well represented on the array with highly redundant and contiguous fosmid clone coverage (Figure 3.6).



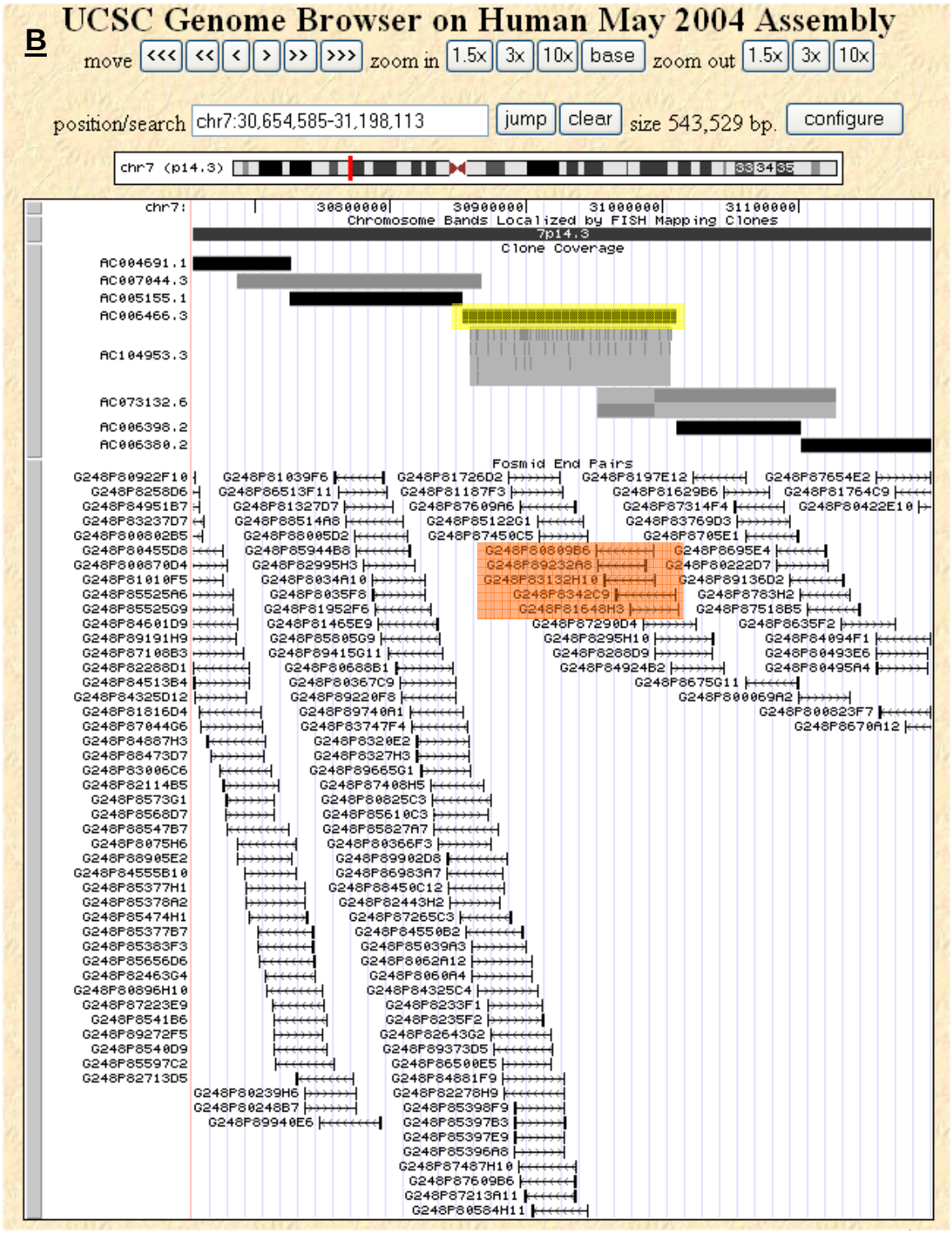


Figure 3.6 ***UCSC view showing fosmid coverage across A chromosome 2 and B chromosome 7 regions of interest. The spanning BAC for chromosome 2 was RP11-680O16, the chromosome 7 spanning BAC clone was CTA-471E18. No BAC end sequence was available for either spanning clone. Spanning BAC clones are highlighted in yellow. Spanning fosmid clones identified by the fosmid array are highlighted in orange.***

Array painting analysis showed a clear transition from high to low ratios for both chromosome 2 and 7 fosmid clones, indicating a simple translocation between the two chromosomes (Figure 3.7).

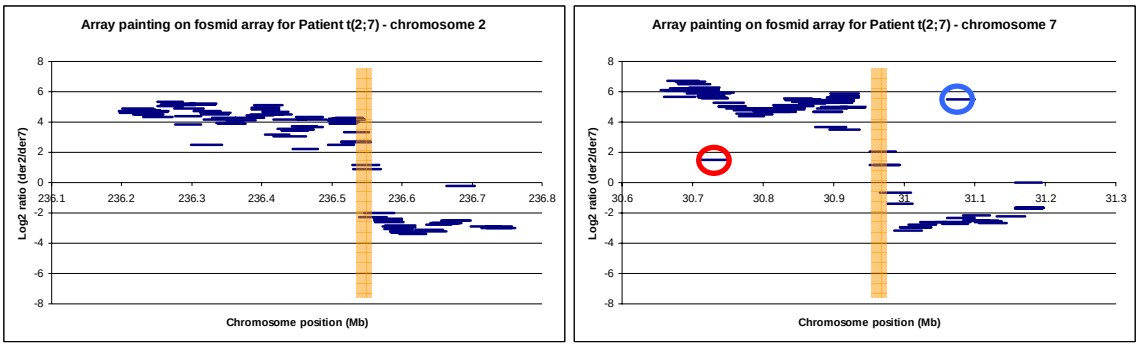


Figure 3.7 Array painting analysis of patient t(2;7)(q37.3;p15.1) using the custom-made fosmid microarray. G248P87223E9 G248P8675G11. The approximate breakpoint region is highlighted in orange.

Analysis of the chromosome 2 clones identified two spanning fosmids; G248P80408F8 at 236,528,698 to 236,567,965bp and G248P84245G4 at 236,530,153 to 236,569,814bp. This reduced the identified chromosome 2 breakpoint region from 150Kb to less than 38Kb. Array painting analysis of the chromosome 7 region identified 5 potential spanning clones; G248P89232A8, G248P80809B6, G248P83132H10, G248P8342C9 and G248P81648H3. Array painting has previously been shown to be a quantitative method (Fiegler et al. 2003b), therefore, the breakpoint containing region could be narrowed down to the 12Kb minimum overlap region from 30,975,849 to 30,987,680bp covered by all of these clones.

Figure 3.7B showed two fosmid clones in regions of good clone coverage with unexpected ratios; G248P87223E9 and G248P8675G11. Subsequent analysis by FISH to verify their chromosomal location showed that G248P87223E9 hybridised to 20q11.21-20q11.23 and G248P8675G11 to 2q12-2q14.2 (Figure 3.8) which is proximal to the chromosome 2 translocation breakpoint proving that these clones had been mis-picked during the production of the microarray.

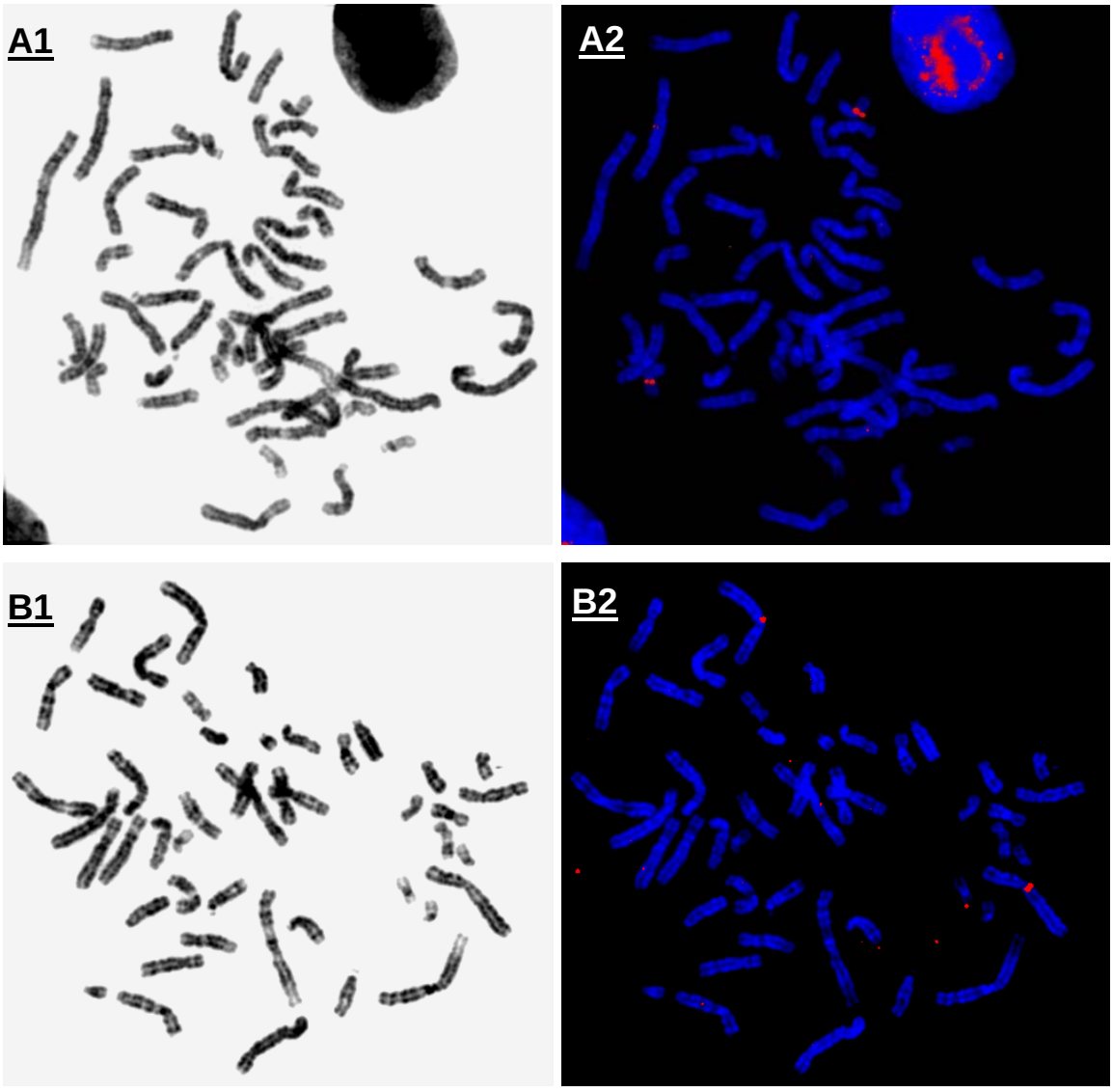


Figure 3.8 FISH results for clones showing unexpected ratios in the patient t(2;7)(q37.3;p15.1) array painting hybridisation; 1 DAPI banding of metaphase chromosomes, 2 metaphase FISH image. A G248P97223E9 at 20q11.21-20q11.23 and B G248P8675G11 at 2q12-2q14.2.

Utilising the $\log_2$ ratios of the clones spanning a chromosome breakpoint it was possible to predict the breakpoint location to within a few kilobases. Excel interpolation analysis of the array data (using the midpoint and $\log_2$ ratio for each spanning clone) predicted the chromosome 2 breakpoint to be at 236,556,071bp and the chromosome 7 breakpoint to be at 30,984,039bp (Figure 3.9).

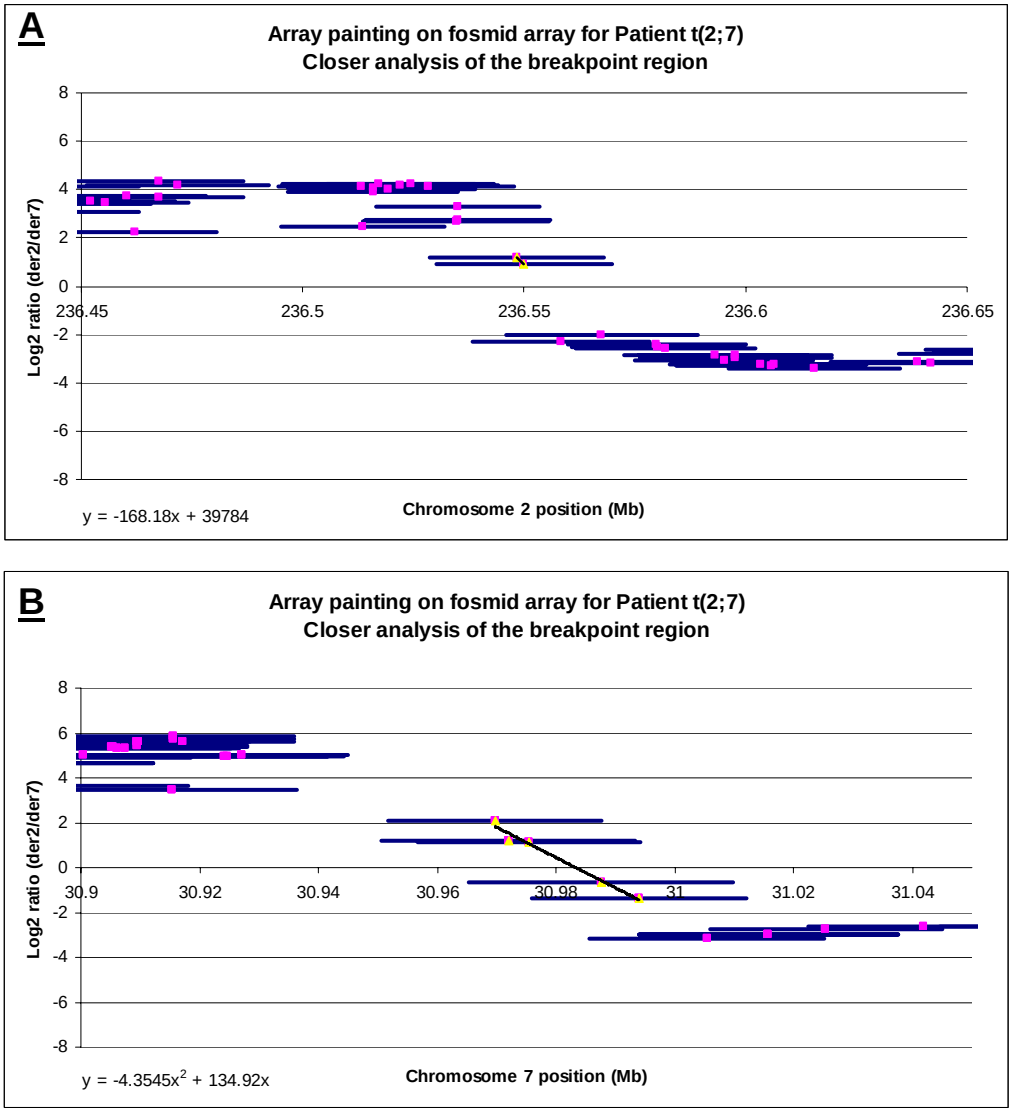


Figure 3.9 Analysis of the array data around the breakpoints for A chromosome 2 and B chromosome 7 for patient t(2;7)(q37.3;p15.1). The pink spots mark the midpoint of each clone.

3.3.2.2 Patient t(3;11)(q21;q12)

For patient t(3;11)(q21;q12) only chromosome 11 clones were included on the array as the chromosome 3 breakpoint had been refined to 2.2Kb previously by STS PCR analysis.

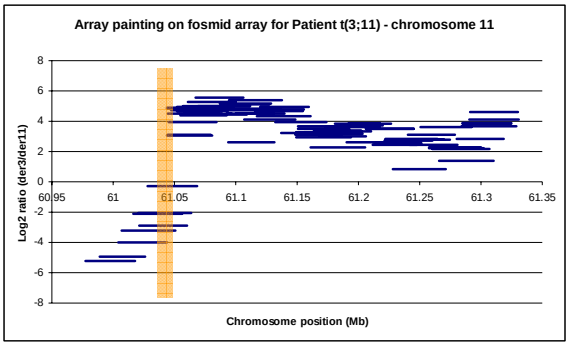


Figure 3.10 Chromosome 11 array painting results for patient t(3;11)(q21;q12) on the custom-made fosmid microarray. The breakpoint containing region is highlighted in orange.

Array painting analysis mapped the chromosome 11 breakpoint to a 23Kb region from 61,027,847 to 61,050,897bp contained within the overlapping clones which showed intermediate ratios on the array; G248P88450F7, G248P80771G8, G248P81609G9, G248P8917A11, G248P87074D8 and G248P82315G10 as highlighted in Figure 3.10 and Figure 3.11.

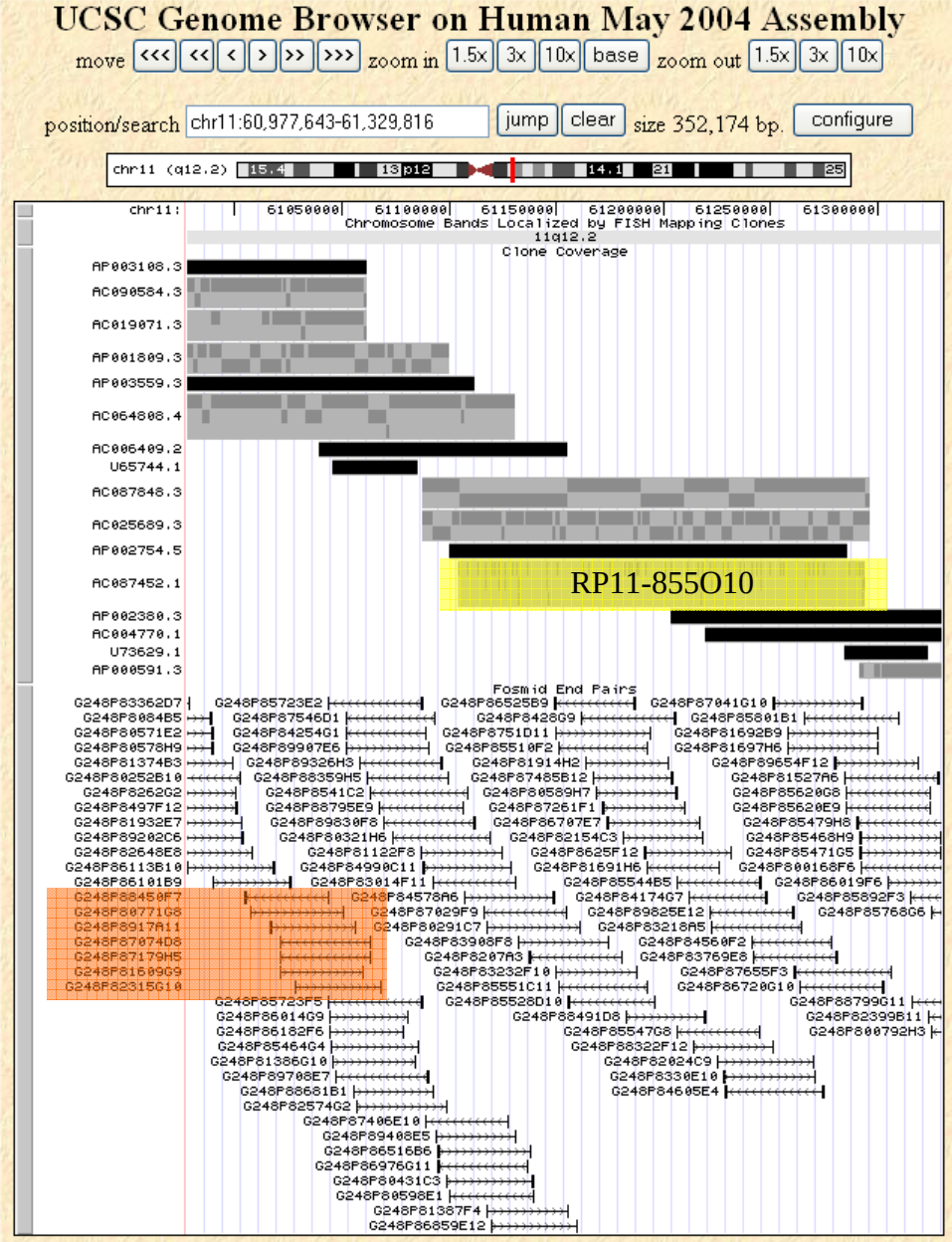


Figure 3.11 UCSC page showing fosmid coverage across the chromosome 11 breakpoint spanning region for patient t(3;11)(q21;q12). The accessioned sequence for the spanning BAC RP11-855O10 is highlighted in yellow. No BAC end sequence was available for this clone. The spanning fosmids are highlighted in orange.

The apparent discrepancy between the spanning BAC clone and the spanning fosmid clones is due to the lack of end sequence data for RP11-855O10. Only the accessioned sequence is represented in Figure 3.11. The full insert length is likely to be much longer.

3.3.2.3 Patient t(7;13)(q31.3;q21.3)

Full fosmid clone coverage was not available for either of the breakpoint regions in patient t(7;13)(q31.3;q21.3). For the genomic regions identified by the spanning BACs, there was a gap of 49Kb in the chromosome 7 fosmid map from 121,232,180 to 121,281,674bp and a gap of 38Kb in the chromosome 13 map from 71,022,384 to 71,060,273bp (Figure 3.13).

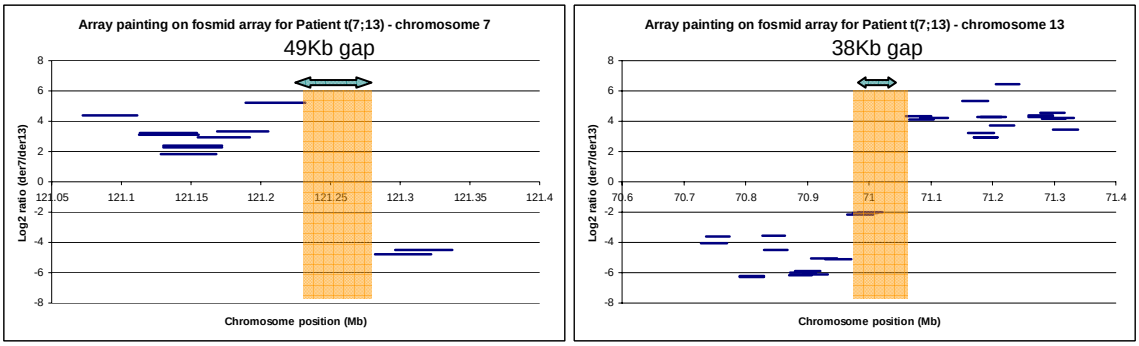
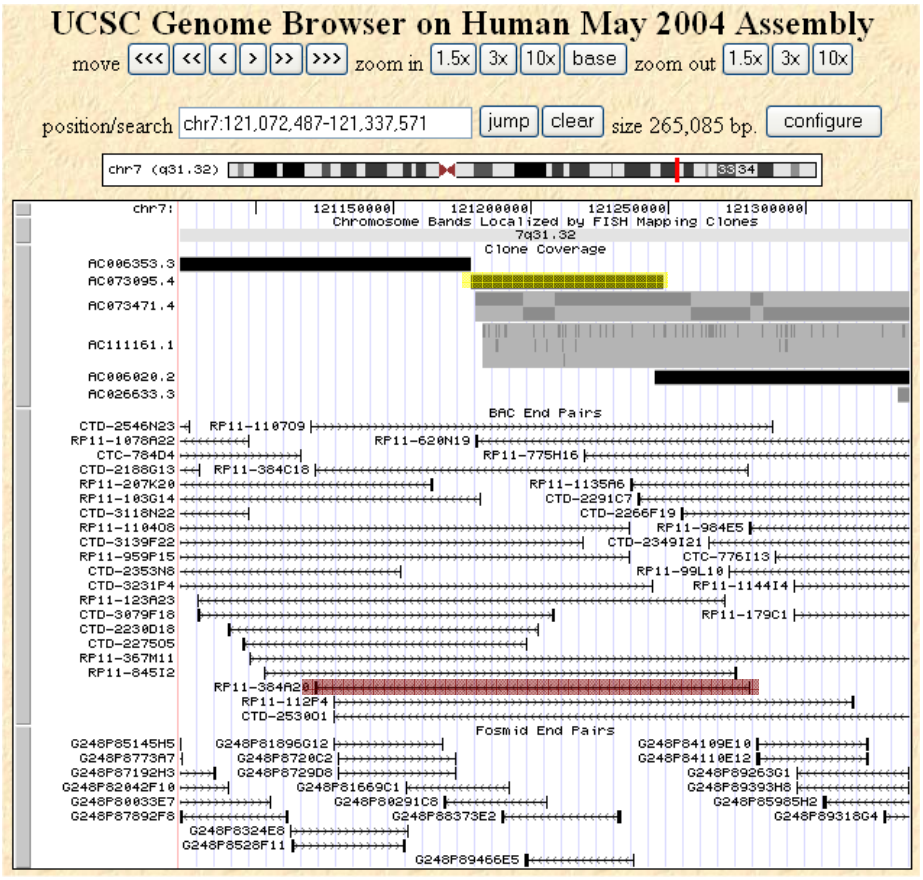


Figure 3.12 Array painting results for patient t(7;13)(q31.3;q21.3) on the custom-made fosmid microarray. Orange arrows depict the approximate breakpoint regions identified.

Array painting (Figure 3.12) refined the chromosome 7 breakpoint to between clones G248P88373E2 and G248P84109E10 corresponding to the 38Kb gap in the fosmid coverage. The chromosome 13 breakpoint appeared to lie towards the distal end of the clones showing an intermediate ratio of approximately -2; G248P88693H1, G248P83744B9 and G248P85941D11. These fosmids lie next to a 38Kb gap from 71,022,389 to 71,060,273bp. Although there were gaps in the fosmid coverage for this region, the spanning BAC clone had been fully sequenced so sequence was available for STS PCR mapping.

A



B

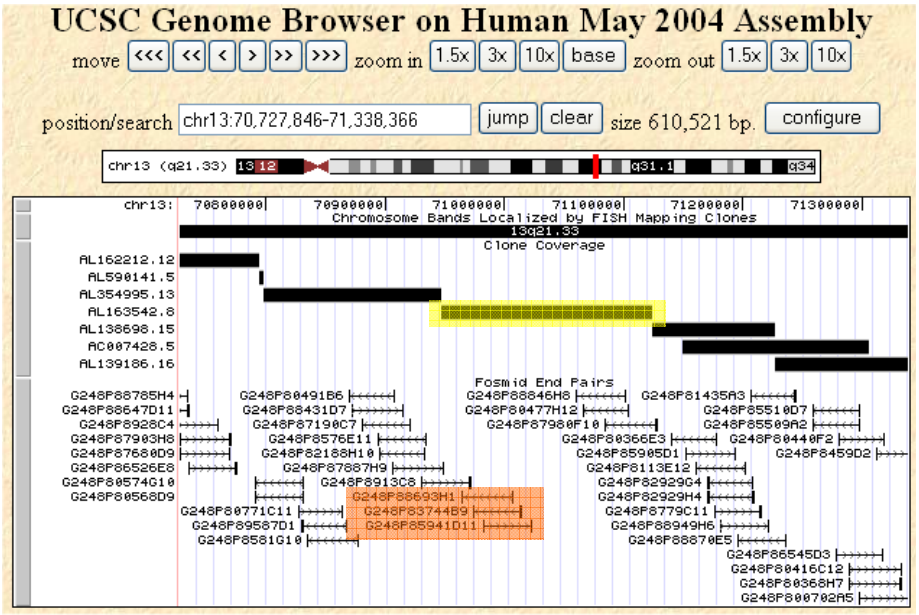
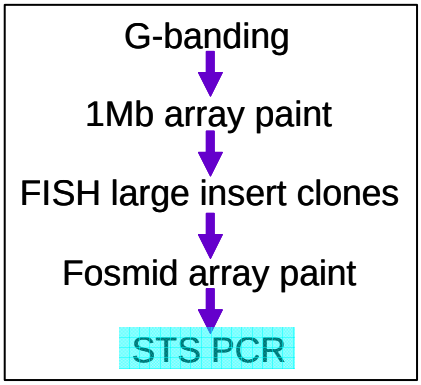


Figure 3.13 ***UCSC download for A chromosome 7 and B chromosome 13 regions of interest for patient t(7;13)(q31.3;q21.3). Spanning BAC clones are highlighted in yellow, and spanning fosmids from the array paint experiment are highlighted in orange. A Spanning BAC RP11-384A20 has been end sequenced and is highlighted in red. B No BAC end sequence was available for RP11-360I23.***

3.4 Further refinement of chromosome translocation breakpoints by STS PCR, PCR microarrays and oligonucleotide microarrays

3.4.1 Refinement of breakpoints by STS PCR



Previous array painting experiments delineated the translocation breakpoints to genomic regions varying from 12-49Kb. Breakpoints were further refined by STS PCR analysis to enable the direct amplification of junction fragments by LR PCR. Genomic regions of interest known to contain the chromosome breakpoints were extended both proximally and distally by a further 3Kb and the sequence exported from NCBI Build 35 for STS PCR primer design (see methods for primer parameters and Appendix A3 for primer sequence).

3.4.1.1 Patient t(2;7)(q37.3;p15.1)

STS PCR primers spaced at 10Kb intervals were designed to cover the chromosome 2 and 7 spanning fosmids. The first round of STS PCR refined the chromosome 2 breakpoint to a 10Kb region from 236,557,221 to 236,567,273bp between primer pair 4 which gave a weak band on derivative 2 and primer pair 5 which gave a strong band on derivative 7 (Figure 3.14).

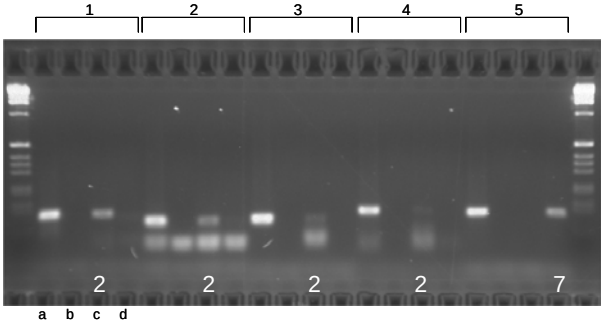


Figure 3.14 Chromosome 2 STS PCR for patient t(2;7)(q37.3;p15.1) at 10Kb resolution resolved the breakpoint to between primers 4 and 5. Templates used; a) whole genomic DNA from a reference sample as a positive control, b) water as a negative control, c) DOP amplified flow-sorted derivative 2, d) DOP amplified flow-sorted derivative 7.

In a subsequent round of PCR using primers designed at 1Kb resolution within this region, only 2 primers gave informative results; both on the derivative 7 (Figure 3.15).

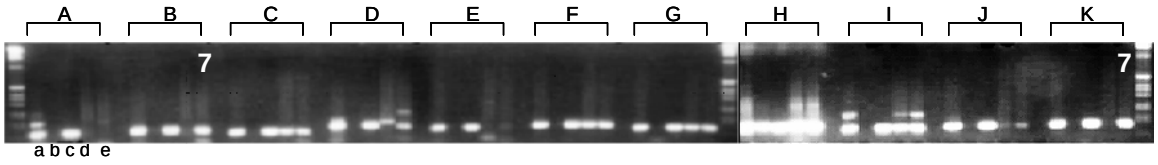


Figure 3.15 Chromosome 2 STS PCR results for patient t(2;7)(q37.3;p15.1) at 1Kb resolution. All informative primers gave products on the derivative 7. Templates used; a) whole genomic DNA from a reference sample as a positive control, b) water as a negative control, c) chromosome 2 spanning BAC clone DNA, d) DOP amplified flow-sorted derivative 2, e) DOP amplified flow-sorted derivative 7.

A further set of primers were designed every 1Kb extending proximal to primer pair 4 to primer pair 3. These primer pairs (L to U) showed a transition from positive PCR products obtained on the derivative 2 (M) to the derivative 7 (Q) (Figure 3.16) which indicated that the breakpoint was between primer pairs M and Q (236,548,584 to 236,553,794bp). Primer pairs N, O and P failed to amplify a product on either derivative chromosome template DNA.

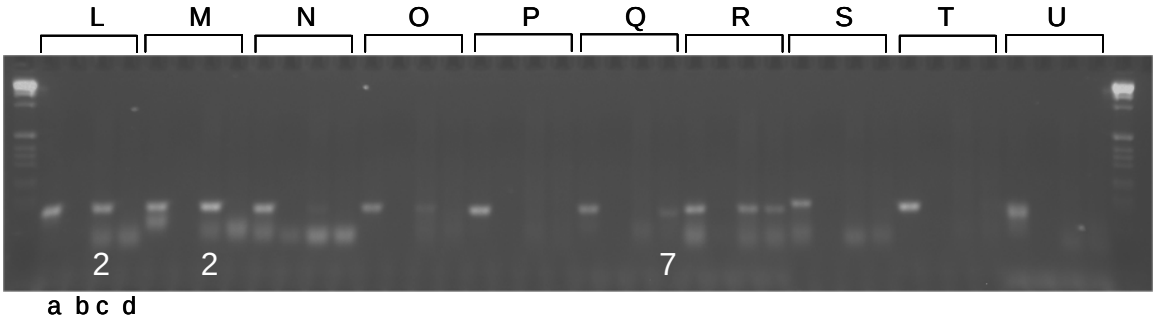


Figure 3.16 Chromosome 2 STS PCR results at 1Kb resolution for patient t(2;7)(q37.3;p15.1) refined the breakpoint region to between primers M and Q. Templates used; a) whole genomic DNA from a reference sample as a positive control, b) water as a negative control, c) DOP amplified flow-sorted derivative 2, d) DOP amplified flow-sorted derivative 7 material.

This data suggested that the weak primer pair 4 result obtained during the 10Kb resolution STS PCR was misleading, and that the breakpoint actually lay between primer pairs 3 and 5. This second round of PCR at 1Kb resolution narrowed the breakpoint to a 5,210bp region from 236,548,584 to 236,553,794bp.

For the chromosome 7 breakpoint, an initial round of STS PCR using primers designed at a 10Kb resolution showed the breakpoint was within a 19,294bp region from 30,979,917 to 30,999,211bp between primers 2 and 4 (Figure 3.17). This region is double the theoretical region that should be resolved using primers at 10Kb resolution as the primer pair in between (primer 3) failed to give a clean PCR result.

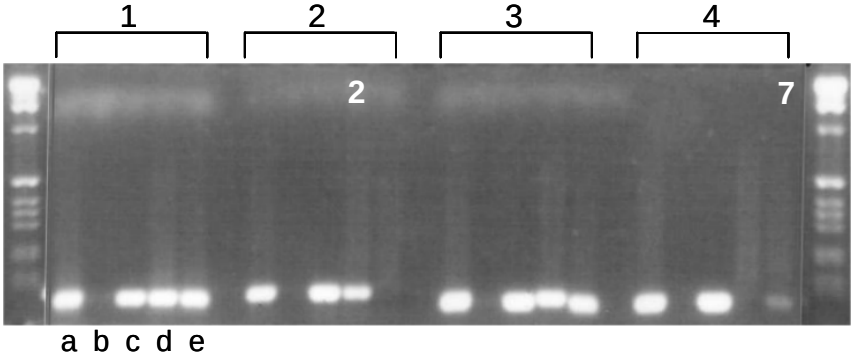


Figure 3.17 Chromosome 7 STS PCR results for patient t(2;7)(q37.3;p15.1) using primers spaced approximately 10Kb apart run with 1Kb marker. Templates used; a) whole genomic DNA from a reference sample as a positive control, b) water as a negative control, c) Chromosome 7 spanning BAC clone, d) DOP amplified flow-sorted derivative 2 and e) DOP amplified flow-sorted derivative 7.

A second round of PCR using primers spaced at approximately 1Kb intervals (Figure 3.18) only managed to refine the breakpoint region to 17,251bp from 30,981,960 to 30,999,211bp between primer B and primer 4 from the initial round of PCR. Half of the primer pairs (D,E,H,I,J,M,O,Q,R) designed across this ~20Kb region produced products on both the derivative 2 and derivative 7 template DNA. Analysis of the 20Kb region of chromosome 7 used to design primers revealed large regions of homology with other regions of chromosome 7 and

regions along chromosome 2 which may explain these double positive results (see Discussion chapter for further analysis of this region).

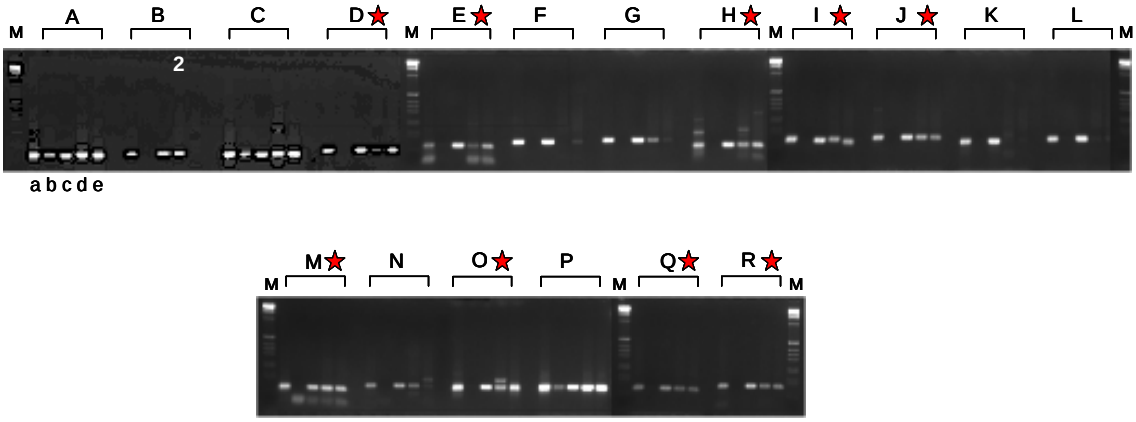


Figure 3.18 Chromosome 7 STS PCR results for patient t(2;7)(q37.3;p15.1) using primers spaced approximately 1Kb apart between primers 2 and 4 from Figure 3.17 run with 1Kb marker. Templates used; a) whole genomic DNA from a reference sample as a positive control, b) water as a negative control, c) Chromosome 7 spanning BAC clone, d) DOP amplified flow-sorted derivative 2 and e) DOP amplified flow-sorted derivative 7.

Alternative strategies of array painting on a PCR product array (3.4.2) and an oligonucleotide array (3.4.3) were investigated to refine the breakpoint region to a level sufficient for LR PCR.

3.4.1.2 Patient t(3;11)(q21;q12)

The chromosome 3 translocation breakpoint had been previously investigated prior to the start of this project. Briefly; STS PCR mapping across the spanning BAC clone sequence at approximately 30Kb resolution mapped the translocation breakpoint to a 40Kb region from 130,728,241bp to 130,767,994bp on chromosome 3. A subsequent round of PCR using primers spaced approximately every 2Kb across this region refined the breakpoint to 2.2Kb from 130,754,507bp to 130,756,748bp.

For chromosome 11, STS PCR mapping using primers spaced at approximately 10Kb intervals mapped the breakpoint to a 20,422bp region from 61,020,792 to 61,041,214bp between primers 1 and 3 (Figure 3.19).

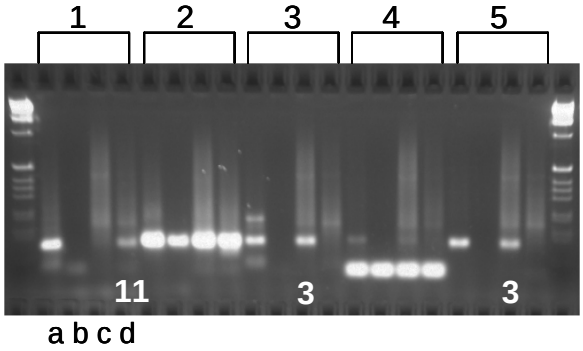


Figure 3.19 Chromosome 11 STS PCR results for patient t(3;11)(q21;q12) at 10Kb resolution resolved the breakpoint to between primers 1 and 3. Templates used; a) whole genomic DNA from reference individual as positive control, b) water as negative control, c) DOP amplified flow sorted derivative 3 material, d) DOP amplified flow sorted derivative 11 material.

A subsequent round of PCR mapping using primers at approximately 2Kb intervals refined the breakpoint region to 7,285bp from 61,026,849 to 61,034,134bp on chromosome 11 between primers C and E (Figure 3.20).

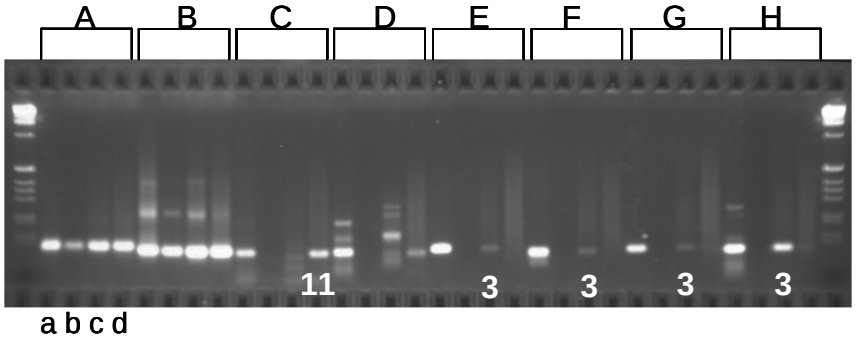


Figure 3.20 Chromosome 11 STS PCR results for patient t(3;11)(q21;q12) at 1Kb resolution refined the breakpoint region to between primers C and E. Templates used; a) whole genomic DNA from reference individual as positive control, b) water as negative control, c) DOP amplified flow sorted derivative 3 material, d) DOP amplified flow sorted derivative 11 material.

3.4.1.3 Patient t(7;13)(q31.3;q21.3)

An initial round of STS PCR at 10Kb resolution mapped the chromosome 7 breakpoint to a 19,851bp region from 121,238,076 to 121,257,927bp between primer 1 which gave a strong band on derivative 7 and primer 3 which gave a weaker band on derivative 13 (Figure 3.21).

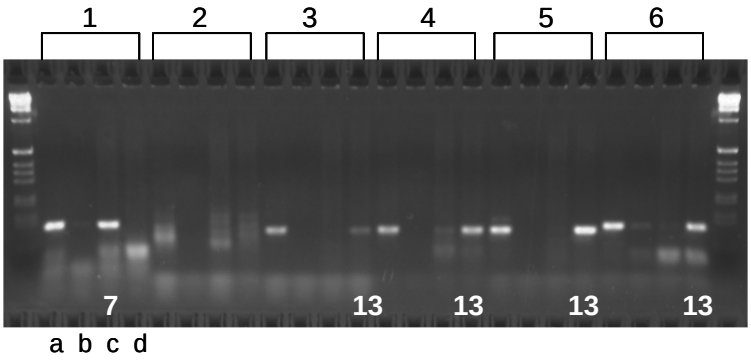


Figure 3.21 Chromosome 7 STS PCR results for patient t(7;13)(q31.3;q21.3) at 10Kb resolution resolved the breakpoint to between primers 1 and 3. Templates used; a) whole genomic DNA from reference individual as positive control, b) water as negative control, c) DOP amplified flow-sorted derivative 7 material, d) DOP amplified flow sorted derivative 13 material.

A subsequent round of PCR refined the breakpoint region to 3,345bp from 121,242,659 to 121,246,004bp (Figure 3.22).

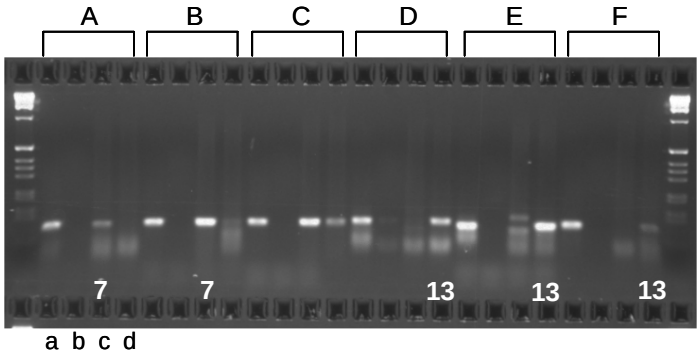


Figure 3.22 Chromosome 7 STS PCR results for patient t(7;13)(q31.3;q21.3) at 1Kb resolution refined the breakpoint region to between primers B and D. Templates used; a) whole genomic DNA from reference individual as positive control, b) water as negative control, c) DOP amplified flow sorted derivative 7 material, d) DOP amplified flow sorted derivative 13 material.

For the chromosome 13 breakpoint, the initial round of STS PCR mapped the breakpoint to a 7,199bp interval between primer pair 2 at 71,016,957bp and primer pair 3 at 71,024,156bp (Figure 3.23).

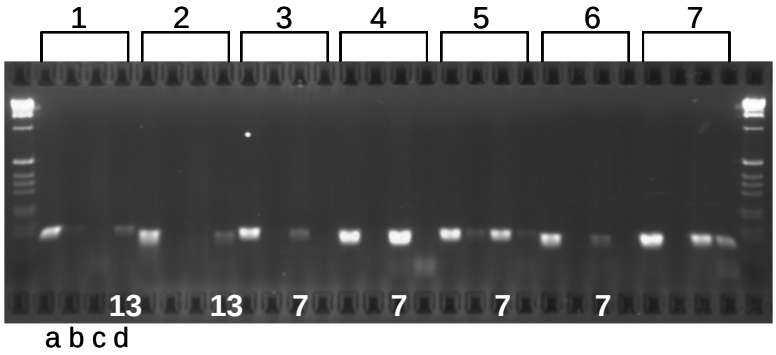


Figure 3.23 Chromosome 13 STS PCR results for patient t(7;13)(q31.3;q21.3) at 10Kb resolution resolved the breakpoint to between primers 2 and 3. Templates used; a) whole genomic DNA from reference individual as positive control, b) water as negative control, c) DOP amplified flow sorted derivative 7 material, d) DOP amplified flow sorted derivative 13 material.

A subsequent round of PCR mapping refined the breakpoint to a 334bp region between primer pairs F and G at 71,023,629 and 71,023,963bp respectively (Figure 3.24).

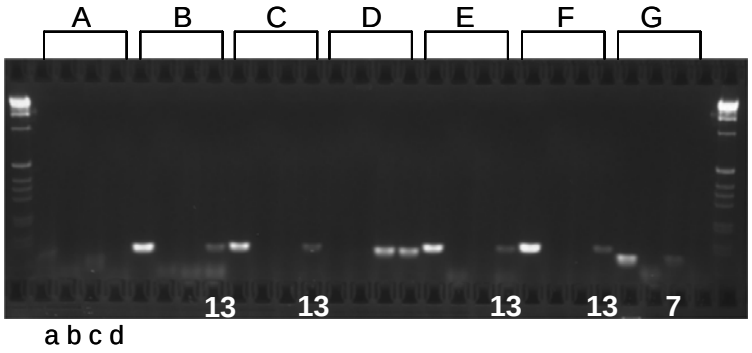


Figure 3.24 Chromosome 13 STS PCR results for patient t(7;13)(q31.3;q21.3). a Initial PCR resolved the breakpoint to between primers 2 and 3. b Subsequent mapping refined the breakpoint region to between primers F and G. Templates used; a) whole genomic DNA from reference individual as positive control, b) water as negative control, c) DOP amplified flow sorted derivative 7 material, d) DOP amplified flow sorted derivative 13 material.

3.4.1.4 Summary of STS PCR data

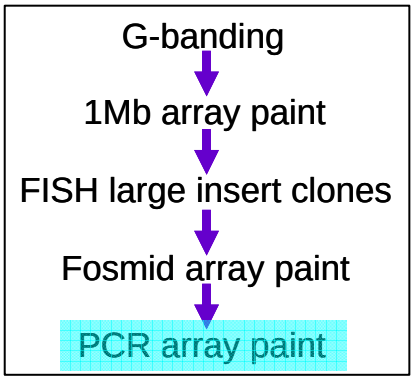
Table 3.2 summarises the mapping of breakpoints by STS PCR. LR PCR can routinely amplify products of up to 10Kb, so with the exception of patient t(2;7)(q37.3;p15.1) all patients had been mapped by STS PCR to a resolution sufficient for junction fragment amplification.

Patient	Chromosome	Fosmid microarray	STS PCR
t(2;7)(q37.3;p15.1)	Chr 2	38Kb	5,210bp
	Chr 7	12Kb	17,251bp
t(3;11)(q21;q12)	Chr 3	n/a	2,241bp
	Chr 11	23Kb	7,285bp
t(7;13)(q31.3;q21.3)	Chr 7	49Kb	3,345bp
	Chr 13	35Kb	334bp

Table 3.2 Summary of breakpoint regions after array painting onto the fosmid microarray and STS PCR.

In virtually all the translocations studied, the breakpoints were refined by STS PCR. However, for patient t(2;7)(q37.3;p15.1) the chromosome 7 breakpoint was refined to 12Kb by fosmid array analysis but only to 17Kb by STS PCR mapping. All primer pairs in this 17Kb region failed to amplify a specific product on either derivative chromosome. To attempt to further refine this breakpoint, higher resolution microarrays were utilised.

3.4.2 Refinement of breakpoints by custom-made PCR product microarrays



STS PCR analysis failed to refine the chromosome 7 breakpoint in patient t(2;7)(q37.3;p15.1) to a resolution sufficient for amplification of the translocation junction by LR PCR, indicating the need for an alternative strategy. PCR primer pairs were designed to create 1Kb products tiling the 17Kb chromosome 7 breakpoint region as defined by STS PCR analysis (Figure 3.25). These primers were used to amplify products to create a custom PCR product array. Each forward primer was

synthesised with an amino group attached to the 5' end to allow for covalent bonding of the product to the microarray slide. This array was then array painted using derivative chromosomes to map the breakpoint further.

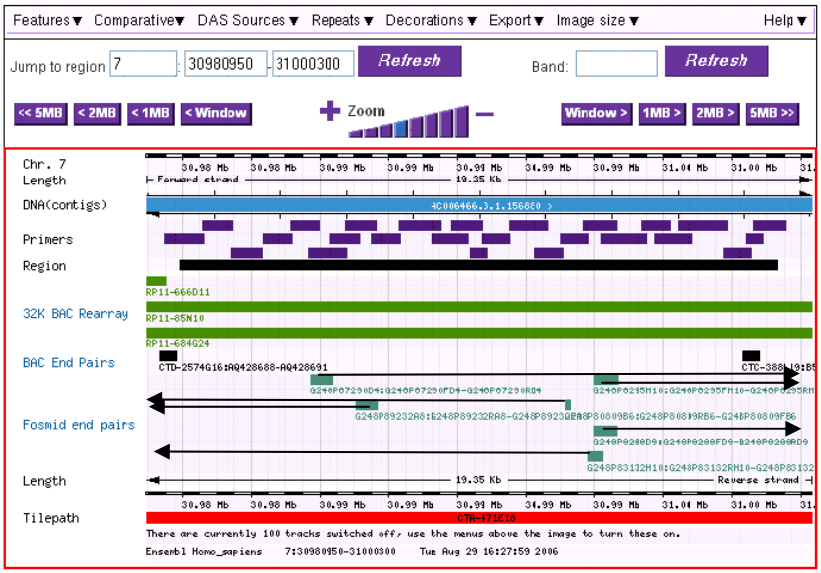


Figure 3.25 Ensembl view showing alignment of PCR products (tiling the 17Kb chromosome 7 region defined by STS PCR analysis) to the human reference sequence. The black arrows on the fosmid end pairs depict the direction of the remainder of the cloned sequence for each fosmid.

Array painting onto the fosmid array (discussed in 3.3.2.1) showed that the fosmid clones depicted in Figure 3.25, G248P89232A8, G248P80809B6 and G248P83132H10 spanned the breakpoint, but that G248P87290D4, G248P8295H10 and G248P8288D9 were distal to the breakpoint.

Array painting onto the PCR product array (Figure 3.26) showed a transition from high to low ratios indicating a shift from one derivative chromosome to the other between a PCR product at 30,982,601 to 30,983,464bp and another at 30,985,159 to 30,985,963bp, refining the breakpoint to 1,695bp. One product (30,983,446 to 30,984,330bp) with an intermediate ratio was observed indicating that this product spanned the breakpoint.

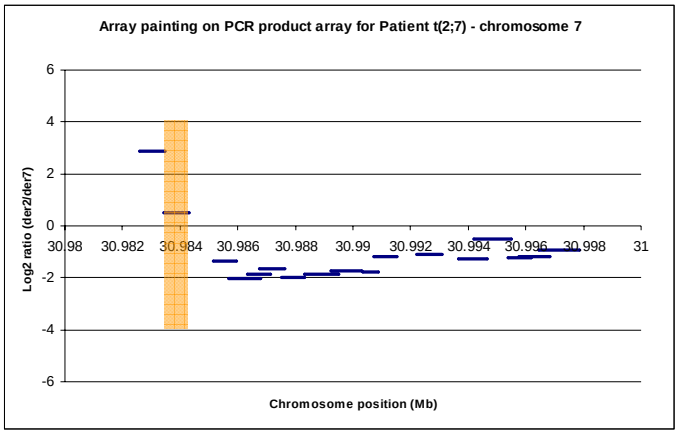
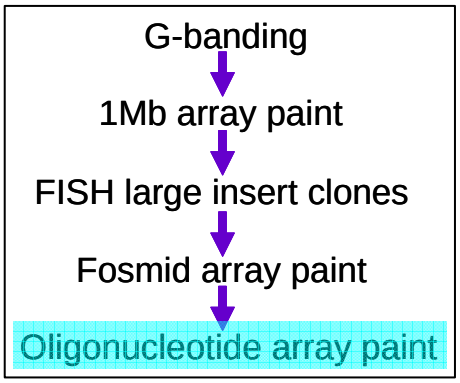


Figure 3.26 Array painting analysis for patient t(2;7)(q37.3;p15.1) on the PCR product array. The breakpoint region is highlighted by the orange region.

3.4.3 Refinement of breakpoints by custom-made oligonucleotide microarrays



A third strategy for refining translocation breakpoints by array painting on oligonucleotide arrays was also investigated for patient t(2;7)(q37.3;p15.1). An ultra-high resolution microarray (NimbleGen Systems, Inc) was constructed using unique, isothermal oligonucleotide probes ranging from 45 to 77bp. The probes were selected from repeat masked sequence spanning the selected regions with 1bp spacing where possible to provide the highest attainable resolution. The chromosome 2 region investigated by the array was from 236,450,000 to 236,650,000bp and the chromosome 7 region from 30,950,000 to 31,300,000bp (Gribble et al. 2006).

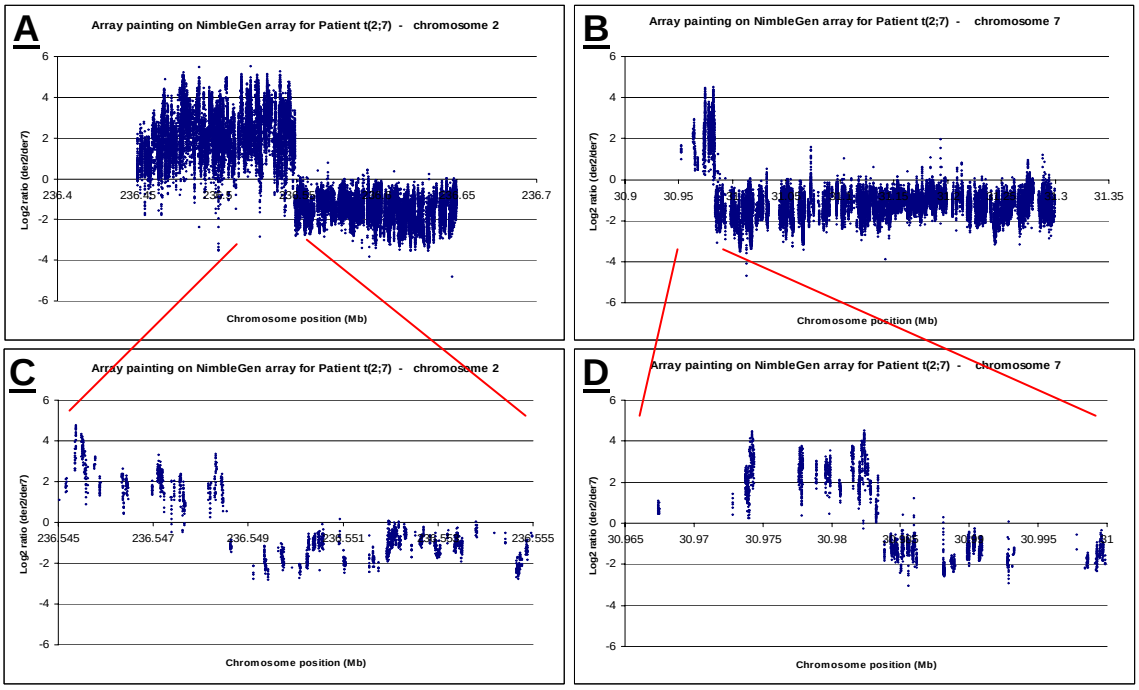
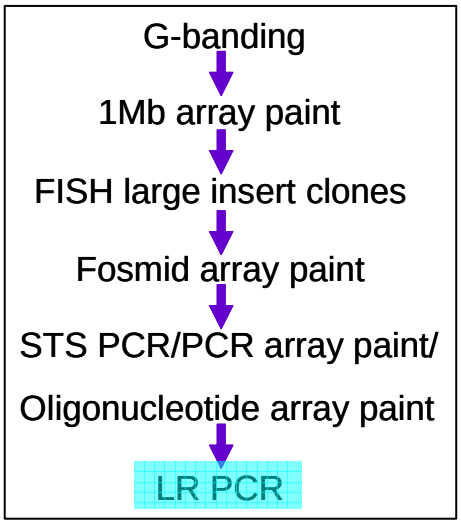


Figure 3.27 Array painting results for patient t(2;7)(q37.3;p15.1) on the NimbleGen oligonucleotide array. A and B show the results for chromosomes 2 and 7. C and D show the breakpoint regions at increased magnitude for chromosomes 2 and 7 respectively.

The ultra-high resolution array painting results (Figure 3.27) showed that the chromosome 2 breakpoint lies between the oligonucleotides at 236,548,539 and 236,548,680bp; a region of 141bp and the chromosome 7 breakpoint lies between the oligonucleotides at 30,983,401 and 30,983,890bp a region of 489bp. This refinement was so precise primers could be designed to amplify the junction fragments directly by PCR.

3.5 Chromosome translocation breakpoint amplification by LR PCR



In all cases, once the breakpoint regions had been refined sufficiently, LR PCR was used to amplify junction fragments.

Patient	Derivative	Primers		Fragment size
t(2;7)	der2	3572_2(2)_1f	ATATACACCATAACAGAAATGACACAGAAG	1.8Kb
		3572_7(2)_2f	GTCAGTCTAAGCATTGAGGTAAAAGACTC	
	der7	3572_7(1)_1r	CCTGCAGATAACACTTAGCTAGAAATAGG	2.0Kb
		3572_2(4)_2r	TAGAAACTGTGCAACTAAAAAGTTTGATAA	
t(3;11)	der3	3606_3(5)_1f	TTTCAAAGGAGTTTAGTTTAAGGATGCTAC	0.9Kb
		3606_11(5)_1r	TCTTACTACCTAGGAGTGTCTGAAGATGAG	
	der11	3606_seq1	ACAGTAACACTGTTTCGAGTTACTACACTG	0.8Kb
		3606_seq2	ATACGTAGCATTAAGCTCATTTCTCAGAC	
t(7;13)	der7	3615_7_2f	AATATAATTCAGATCAAGTGCACTAACTTC	0.7Kb
		3615_13_2r	TAAATCAATAGGGTGACTATAGTTTACAGC	
	der13	3615_13_3f	TTTACAACTAGGTGCAAGAGATTTACTAAC	3.5Kb
		3615_7_2r	GTATATGTAAATTCAGCTACAGAAACAGGT	

Table 3.3 Primers used for amplification of junction fragments for patients t(2;7)(q37.3;p15.1), t(3;11)(q21;q12) and t(7;13)(q31.3;q21.3).

Junction fragments were amplified successfully for all derivative chromosomes (Figure 3.28). Each primer pair (Table 3.3) was tested on genomic DNA from a normal reference individual to ensure that the amplified product was unique to the patient.

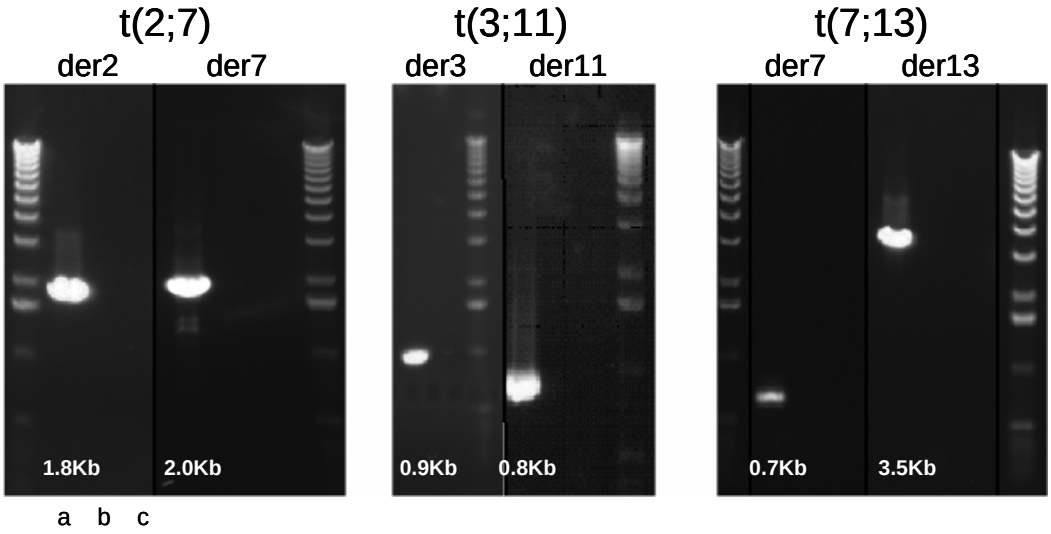
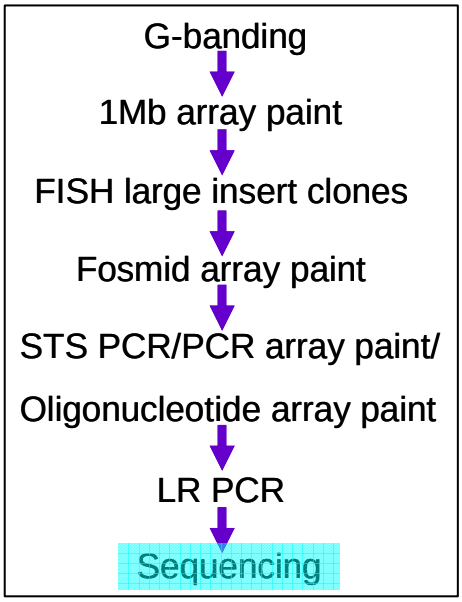


Figure 3.28 Agarose gel analysis of junction fragments for all 3 patients run against 1Kb marker. Templates used; a) patient genomic DNA, b) reference genomic DNA, c) water as negative control.

3.6 Sequence of junction fragments



The same PCR primers used to generate junction fragments were used to sequence the products. Approximately 400-500bp of sequence was obtained from both the forward and reverse primers. A subsequent round of primer design at the end of the sequence read generated the next sequence read, "walking" across the product to generate the sequence necessary. The sequence reads were blasted back against the sequence from NCBI Build 35 to align the derivative sequence to the reference sequence for exact identification of the translocation breakpoints (Figure 3.29).

```
t(2;7)(q37.3;p15.1)
2r/c;  aaatgaaaatcccccaagatcacccaggaagcagccctgggtataagcctcacactcccaacat
chr2;  actgtctgcccatgttgggagtgtgaggcttatacccagggctgcttcctgggtgatcttgggg
der2;  actgtctgcccatgttgggagtgtgaggcttataTTAAGGCTTTAAATATCTGAATGTAGCCAT
der7;  cagcccaggaagcagtttaatctattaatctttgTTAATAGATTAAACAAAAACTGAGCGTAAA
chr7;  GGTACCTCATGGCTACATTCAGATATTTAAAGCCTTAATAGATTAAACAAAAACTGAGCGTAAA
7r/c;  ACCCTGAGTTTACGCTCAGTTTTTGTTTAATCTATTAAGGCTTTAAATATCTGAATGTAGCCAT

t(3;11)(q21;q12)
chr3   acaccctgatcctgagttcactcctcggcccgcccatccaagcgccccgtgcggctggcgtt
der3;  acaccctgatcctgagttcactcctcggcccGGGCTGGATGGGCAGGTAGGGGGCGGGCTCCGG
der11; CGGGGCGGGGAATCTCTCGGCTTGTGCTTGCcccatccaagcgccccgtgcggctggcgttctc
chr11; CGGGGCGGGGAATCTCTCGGCTTGTGCTTGCTCCGCGGTGGGCTGGATGGGCAGGTAGGGGGCG

t(7;13)(q31.3;q21.3)
chr7;  tagtgattcggccttgcatgctacgcctgtatttcccagtgtcgatgtgtcatttgaatccatc
der7;  tagtgattcggccttgcatgctacgcctgGCTTTGTATGAAAATGGTCCATAAGATTGGATGCT
der13; GTGATACATAATGATTTTCTTTCAAAAGGGCTTTttcccagtgtcgatgtgtcatttgaatcca
chr13; GTGATACATAATGATTTTCTTTCAAAAGGGCTTTGTATGAAAATGGTCCATAAGATTGGATGCT
```

Figure 3.29 Sequence obtained across the junction regions for all three patients discussed in this chapter. Highlighted bases are; deleted bases (red), duplicated bases (green), inserted bases (yellow), bases of ambiguous origin which could have come from either donor chromosome (blue).

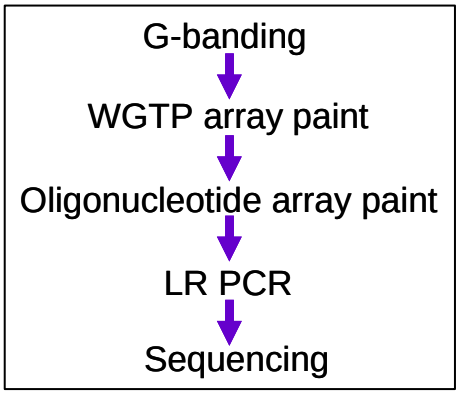
The breakpoint positions for each patient are detailed in Table 3.4. Each breakpoint was accompanied by varying numbers of deleted and/or duplicated bases. For patients t(3;11)(q21;q12) and t(7;13)(q31.3;q21.3) there were homologous bases at the breakpoints present on both donor chromosomes so the breakpoint to the single basepair level could not be determined; patient t(3;11)(q21;q12) had a 2bp homology (GC) and patient t(7;13)(q31.3;q21.3) a single basepair (T) homology.

Patient	Chromosome	Breakpoint position	Additional complexity at the breakpoint	
t(2;7)	2	236,548,579	236,548,580-236,548,586 deleted	19bp insertion on der7
	7	30,983,603	30,983,604-30,983,607 duplicated	
t(3;11)	3	130,755,505	130,755,506 deleted	
			130,755,507-130,755,508 either deleted or on der11	
	11	61,032,619	61,032,620-61,032,621 either deleted or on der11	
			61,032,622-61,032,629 deleted	
t(7;13)	7	121,245,446	121,245,447-121,245,448 deleted	
			121,245,449 either deleted or on der13	
	13	71,023,877	71,023,878-71,023,881 duplicated	
			71,023,822 either deleted or on der13	

Table 3.4 Translocation positions from NCBI Build 35 and additional sequence complexity at the breakpoints for all 3 patients.

Analysis of the sequence obtained from the two junction fragments and normal reference sequence did not reveal any homology at the breakpoint between the donor chromosomes in any of the three patients. Further investigation of the surrounding sequence, gene analysis and possible mechanisms are discussed in detail in Chapter 4.

3.7 Most rapid method of translocation breakpoint mapping



Advances in array technology over the course of this study have enabled the production of a whole genome tile path (WGTP) array consisting of large insert clones which tile the whole genome, resulting in a resolution of approximately 150Kb. A single hybridisation on the Sanger WGTP array can map both breakpoints of a reciprocal translocation within large insert clones and replaces a

more conventional approach of a 1Mb array paint and subsequent FISH hybridisations (Gribble et al. 2005). Once a spanning large insert clone has been identified by array painting using the WGTP array, the genomic region can be investigated by an oligonucleotide array paint analysis (NimbleGen Systems, Inc.). Using this custom-made ultra-high resolution oligonucleotide array as described in 3.4.3 will normally identify the breakpoints at a resolution sufficient for the direct amplification of junction fragments by PCR.

To demonstrate this approach, patient t(2;7)(q37.3;p15.1) was array painted onto the Sanger WGTP array to identify the spanning BAC clones (Figure 3.30). The chromosome 2 spanning clone was identified as RP11-401E12 and the spanning chromosome 7 clone identified as RP11-259J14.

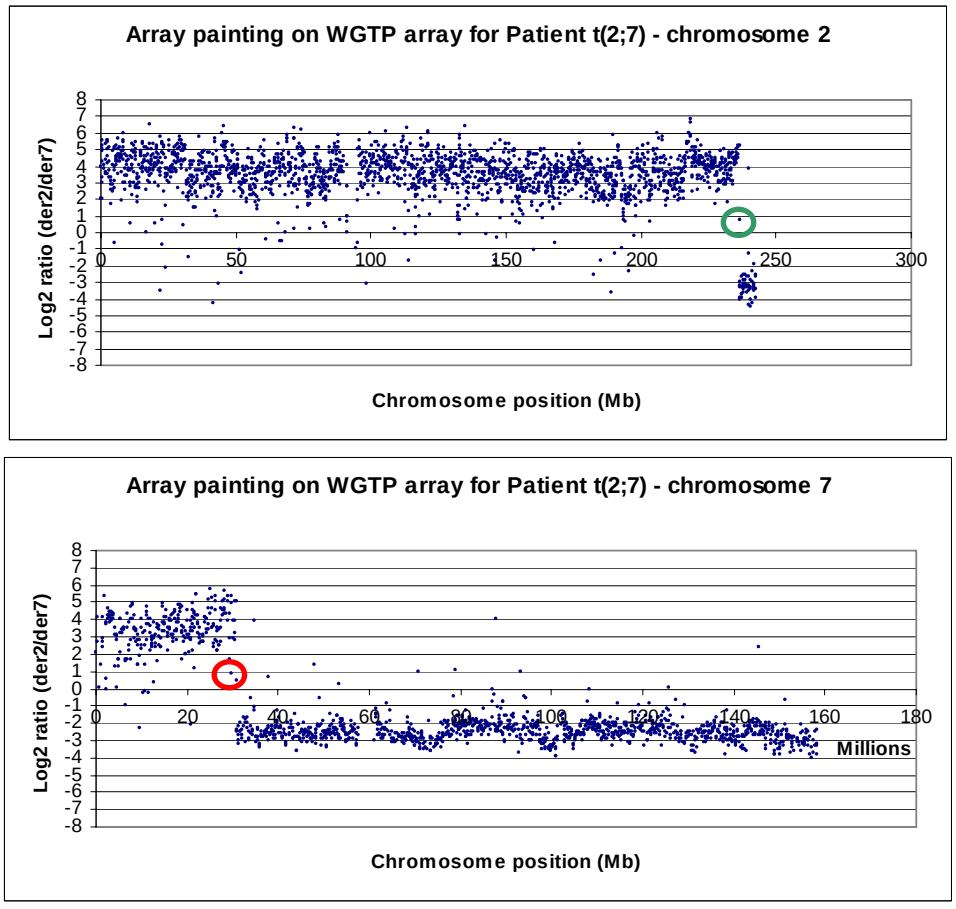


Figure 3.30 Array painting results for patient t(2;7)(q37.3;p15.1) on the whole genome tilepath array. RP11-401E12 and RP11-259J14.

The spanning clones identified by the WGTP array paint were not identical to the clones previously identified by FISH as the clones used for FISH experiments were not selected for generation of the WGTP array (Table 3.5). However, the clones identified by both approaches contain similar and overlapping genomic regions. All clones selected for the Sanger WGTP array were verified by end sequencing.

	Clones found by FISH	Clone found by WGTP array
Chr 2	RP11-680O16	RP11-401E12
	236,390,125 to 236,562,868bp	236,389,125 to 236,551,779bp
Chr 7	CTA-471E18	RP11-259J14
	30,853,334 to 31,010,213bp	30,899,687 to 31,060,782bp

Table 3.5 Comparison of spanning clones for patient t(2;7)(q37.3;p15.1) found by FISH and WGTP array painting.

A degree of variation was seen amongst the clones after hybridisation to derivative chromosomes. This is shown by the outliers in Figure 3.30. This could have arisen because of the repetitive nature of the human genome or from the amplification step used to increase the amount of probe prior to hybridisation to the microarray. It is estimated that approximately 5.4%-50% of the human genome is repeated (Cheung et al. 2001; Lander et al. 2001). Regions of homology within the genome can alter ratios as the probe cross-hybridises to multiple clones on the microarray. Also seen in the data was a wave phenomenon. Amplification of the flow-sorted derivative chromosomes may introduce a level of bias if certain regions of the template DNA are amplified preferentially, resulting in an unequal amount of DNA being generated across the whole template and so hybridising to the microarray causing the wave pattern.

Array painting onto the WGTP array followed by array painting on a custom-made oligonucleotide array is currently the most rapid and efficient approach to breakpoint mapping.

3.8 Investigation of further rearrangement around translocation breakpoints using fosmid microarrays

Previous reports have shown that translocations are sometimes accompanied by additional complexity at or around the breakpoints. A combination of three studies (Kumar et al. 1998; Astbury et al. 2004; Patsalis et al. 2004) using FISH and PCR reported that 8 out of 30 apparently balanced reciprocal translocations are more highly rearranged than identified by G-banding. An initial CGH screen on a 1Mb resolution microarray showed a further imbalance in patient t(2;7)(q37.3;p15.1) on chromosome 3 (Gribble et al. 2005). To investigate any additional complexity close to the breakpoints in each of the 3 patients array CGH was performed on the custom-made fosmid array (Figure 3.31).

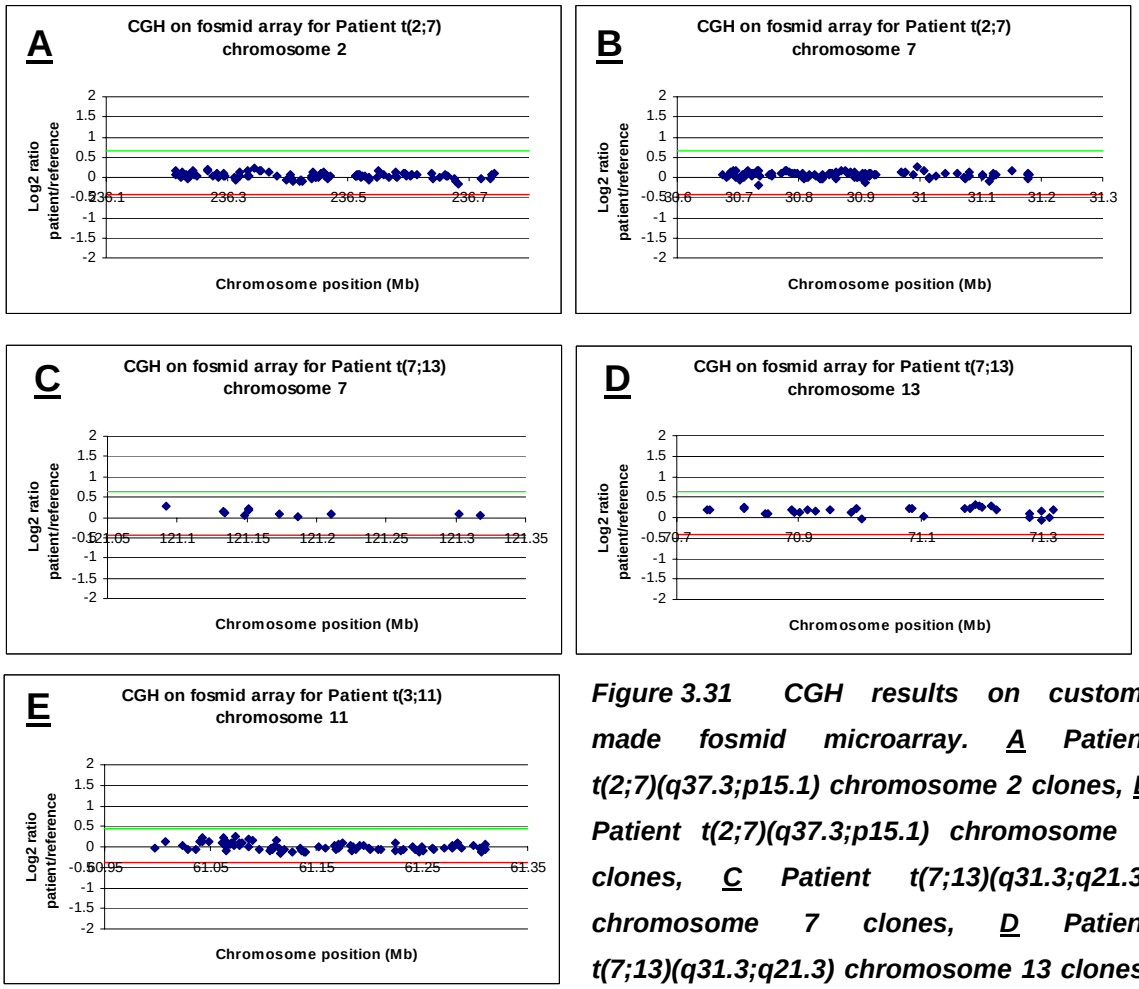


Figure 3.31 CGH results on custom-made fosmid microarray. A Patient t(2;7)(q37.3;p15.1) chromosome 2 clones, B Patient t(2;7)(q37.3;p15.1) chromosome 7 clones, C Patient t(7;13)(q31.3;q21.3) chromosome 7 clones, D Patient t(7;13)(q31.3;q21.3) chromosome 13 clones, E Patient t(3;11)(q21;q12) chromosome 11 clones.

None of the patients showed any additional complexity around the breakpoints at the resolution of the fosmid array.

3.9 Investigation of additional imbalance in patient t(2;7)(q37.3;p15.1)

Whilst no additional complexity was found close to the translocation breakpoints, the initial 1Mb array CGH screen identified a duplication of material at 3p26.3 in patient t(2;7)(q37.3;p15.1) which was confirmed by genotyping to be present in the father and the proband (Gribble et al. 2005). The frequency of interstitial duplications is believed to be approximately 1 in 4000 (Shaffer and Lupski 2000). A possible mechanism for these duplications (and corresponding deletions) is through non-allelic homologous recombination by virtue of homology of the sequence along the affected chromosome (Figure 3.32).

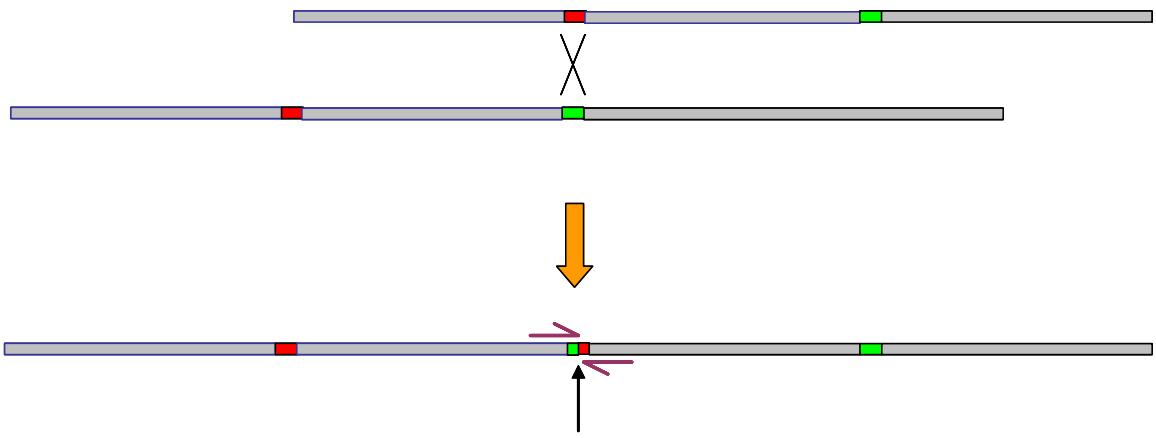


Figure 3.32 Possible mechanism for the 3dup(p26.3)(p26.3) in patient t(2;7)(q37.3;p15.1). The black arrow indicates the duplication junction. The purple arrows show the orientation of primers necessary to amplify the proposed duplication junction.

The two sister chromatids align close to each other during meiosis to undergo homologous recombination. If the chromatids are misaligned, then unequal crossing over can occur resulting in the duplication seen in the patient.

The 3dup(p26.3)(p26.3) identified by the 1Mb array CGH screen was mapped further using a custom-made fosmid microarray and LR PCR.

3.9.1 Delineation of chromosome 3 duplication breakpoints in patient t(2;7)(q37.3;p15.1) by microarray analysis

A high resolution fosmid array was constructed using 497 fosmids at full redundancy selected from NCBI Build 35 to refine the duplication breakpoints initially identified by the 1Mb array CGH screen (Gribble et al. 2005).

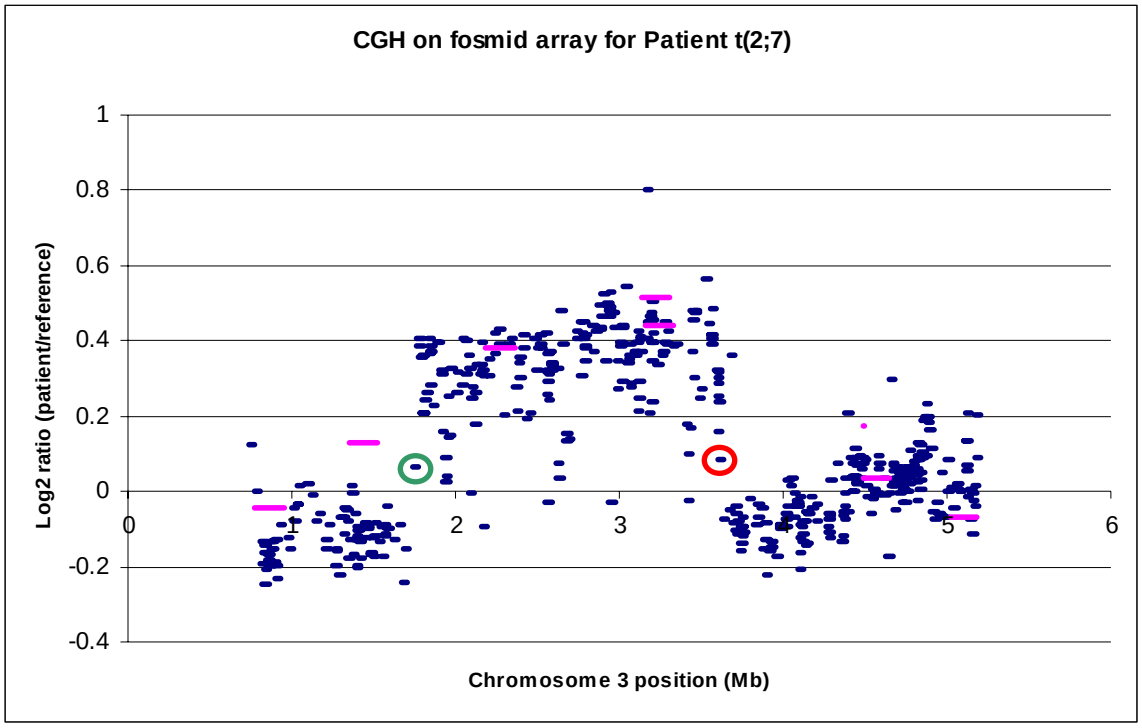


Figure 3.33 CGH results for patient t(2;7)(q37.3;p15.1) on custom-made fosmid microarray. Fosmid clones are in blue, 1Mb clones from original CGH screen (Gribble et al. 2005) in pink. G248P8110H1 G248P89401F1.

Fosmid array CGH was used to identify and refine the duplication breakpoints. CGH using patient DNA and a reference DNA on this microarray identified the distal breakpoint in clone G248P8110H1 and the proximal breakpoint in G248P89401F1 (Figure 3.33). As part of the original investigation into this amplification, genotyping showed that a duplication was also present in the phenotypically normal father at the same location (Gribble et al. 2005). To confirm that the duplication was the same in the patient as it was in the father a

CGH experiment competitively hybridising the patient's DNA against the father's DNA was performed (Figure 3.34).

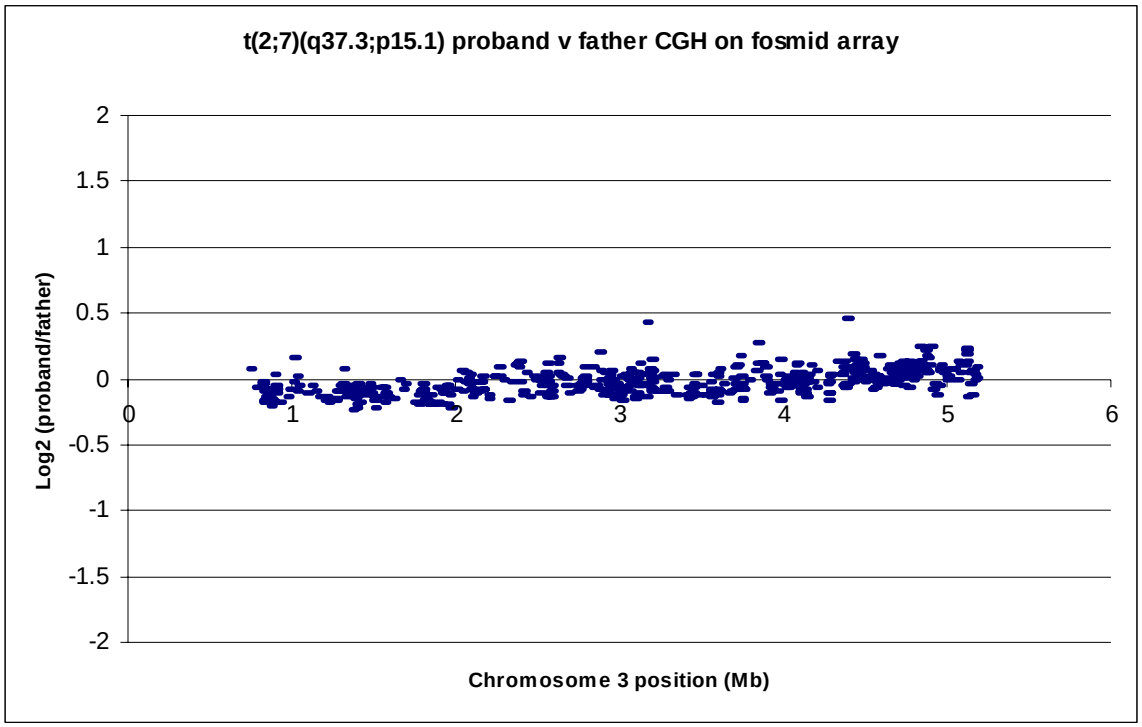


Figure 3.34 CGH results for patient t(2;7)(q37.3;p15.1) versus the patient's father on custom-made fosmid microarray.

No shift in ratios was observed from the CGH experiment for the clones spotted on the microarray, indicating that there was no difference in copy number between the father and the proband verifying that the same duplication was present in both DNA samples.

To further refine the breakpoints an oligonucleotide array was designed for the distal region 1,600,000 to 1,900,000bp and the proximal region 3,500,000 to 3,800,000bp on chromosome 3.

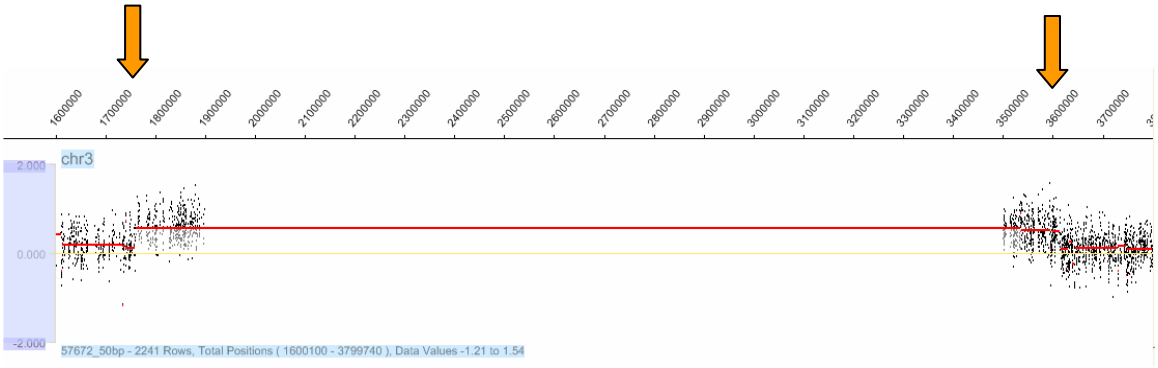


Figure 3.35 CGH for patient t(2;7)(q37.3;p15.1) onto a custom-made oligonucleotide array to refine the chromosome 3 duplication. The red lines show the averaged data over 50bp segments. Breakpoint regions are highlighted by the orange arrows.

Array CGH on the custom-made oligonucleotide array (Figure 3.35) refined the distal breakpoint to within 4,947bp from 1,759,196 to 1,764,153bp and the proximal breakpoint to within 1,105bp from 3,613,988 to 3,615,093bp.

3.9.2 Amplification and sequence of chromosome 3 duplication junction in patient t(2;7)(q37.3;p15.1)

LR PCR with primers designed to the breakpoint regions refined by the oligonucleotide array amplified products using patient t(2;7)(q37.3;p15.1) DNA as template. The same LR PCR primer pairs used on the father's DNA also produced a product of the same size as the proband. No product was obtained using the same primer pairs on the mother's DNA (Figure 3.36). The product obtained using the proband's DNA was sequenced to identify the duplication breakpoints.

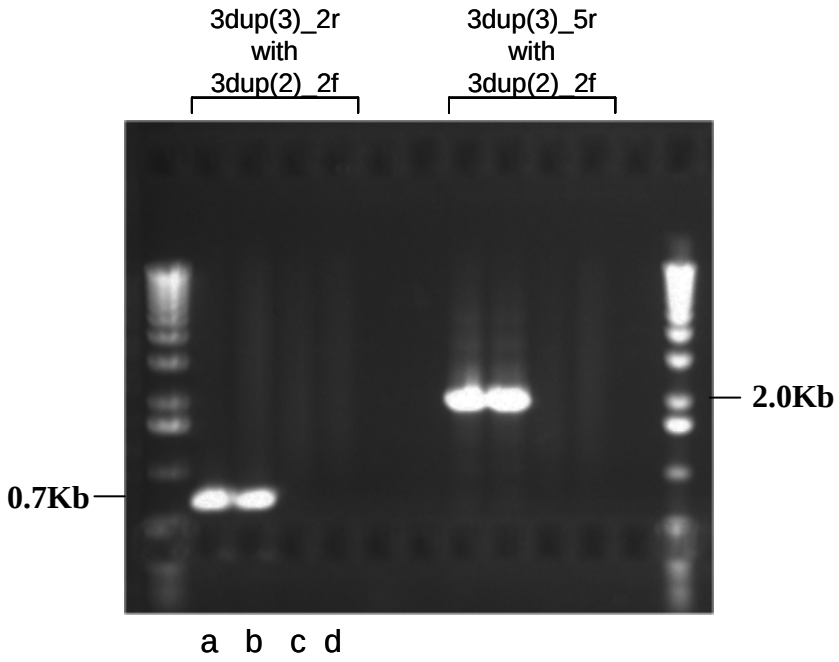


Figure 3.36 Agarose gel analysis of LR PCR results for amplification of chromosome 3 duplication junction in patient t(2;7)(q37.3;p15.1) run with 1Kb marker. Templates used; a) Patient genomic DNA, b) genomic DNA from father, c) genomic DNA from mother, d) water for negative control.

The breakpoint junction fragments were sequenced and the sequence obtained was aligned using ClustalW (http://www.ebi.ac.uk/clustalw/index.html?) and sequence homologies highlighted using the Genedoc software (http://www.psc.edu/biomed/genedoc/). The precise breakpoints within the sequence could not be found as both the proximal and distal breakpoints fall within Alu elements with 97% homology at approximately 1.7 and 3.6Mb along chromosome 3 (Figure 3.37). The distal breakpoint was within a 260bp region from 1,756,993 to 1,757,252bp and the proximal breakpoint was within a 258bp region from 3,614,132 to 3,614,389bp.

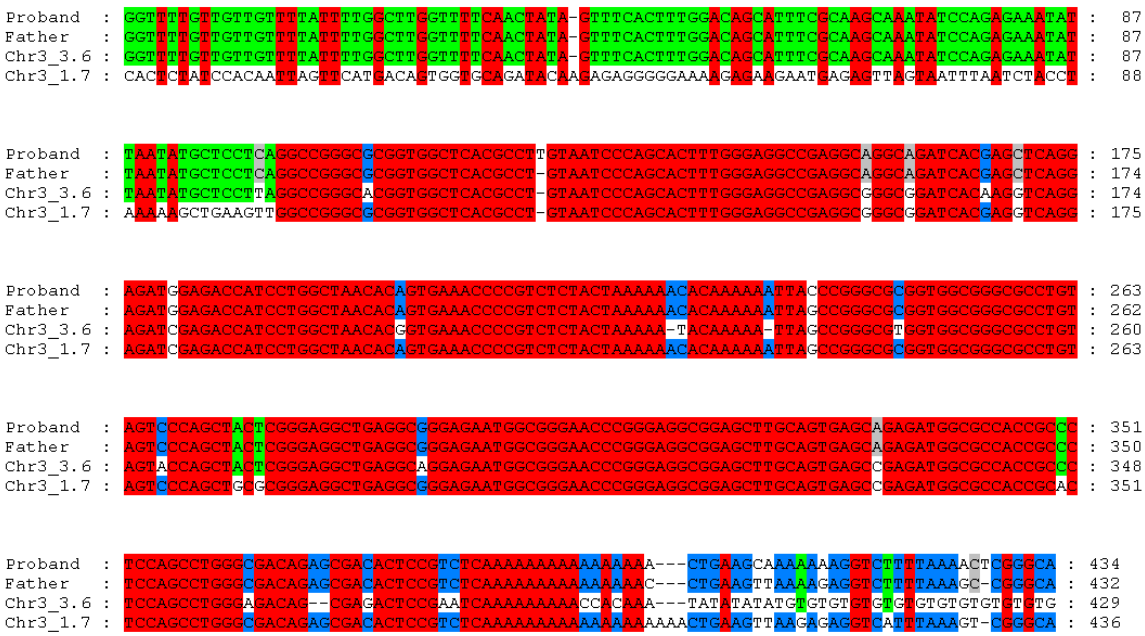


Figure 3.37 Sequence alignment of chromosome 3 duplication junction for patient t(2;7)(q37.3;p15.1), the patient's father, Alu elements at 3.6 and 1.7Mb on chromosome 3 at the rearrangement breakpoints. Shading; patient and father sequence match both Alu sequences (red), patient and father sequence match the Alu sequence at 3.6Mb on chromosome 3 (green), patient and father sequence match the Alu sequence at 1.7Mb on chromosome 3 (blue).

This sequence supports the hypothesis that the duplication arises from unequal crossing over during homologous recombination between sister chromatids mediated by an Alu repeat.

3.10 Conclusions

This chapter describes multiple techniques to refine chromosome breakpoints; Figure 3.38 shows the resolution of each of these techniques.

Simultaneous CGH and array paint hybridisations onto the same microarrays can identify and map regions of copy number loss or gain and chromosome translocation breakpoints.

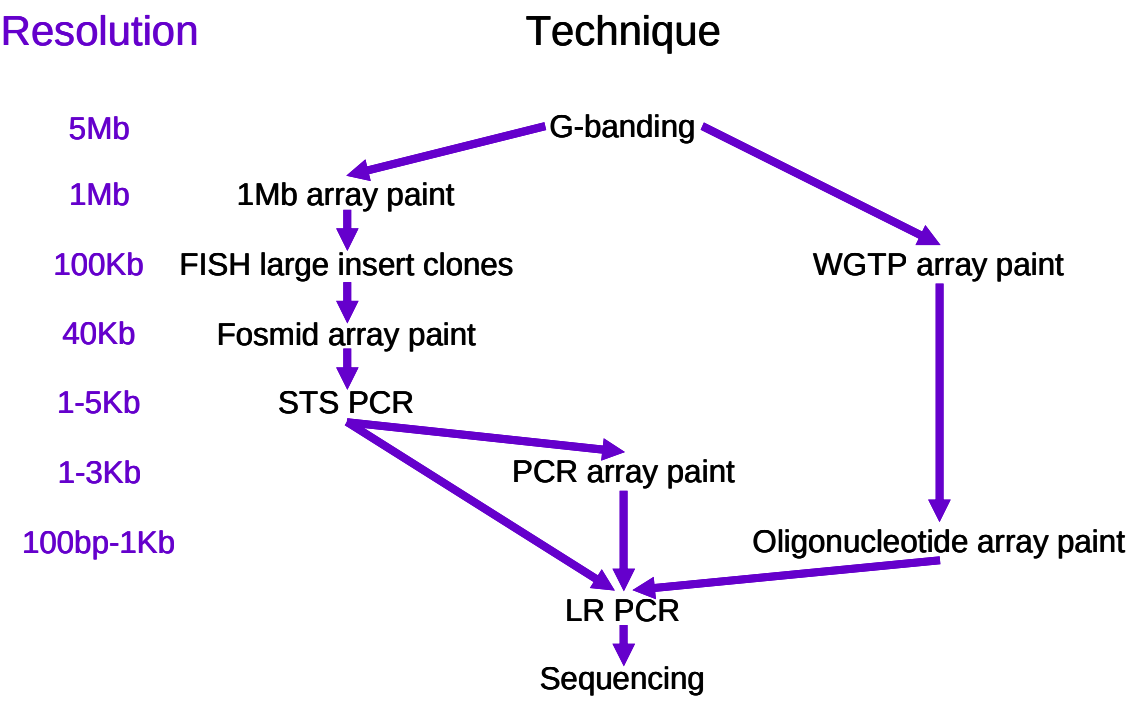


Figure 3.38 Schematic showing techniques used in this chapter to map rearrangement breakpoints.

Once the rearrangement breakpoints were mapped and sequenced the genomic regions spanning and flanking the breakpoints were submitted for bioinformatic analysis to look at the genomic architecture around the breakpoints and gene disruption. This analysis, along with possible mechanisms for the rearrangements are discussed in Chapter 4.

4 Results: Bioinformatic investigation of rearrangement breakpoints

4.1 Introduction

The methods used to obtain the sequence across the translocation junctions in 3 patients with *de novo* constitutional reciprocal translocations have been previously discussed in Chapter 3. This chapter discusses the effect the translocation may have on each patient's phenotype and the possible mechanisms underlying the rearrangements.

Translocation breakpoint positions were originally mapped in NCBI Build 35. The remapped breakpoint positions according to NCBI Build 36 of the human genome are detailed in Table 4.1.

Patient	Chromosome	Breakpoint position	Breakpoint position
		May 2004	**March 2006**
t(2;7)(q37.3;p15.1)	2	236,548,579	236,431,318
	7	30,983,603	31,176,888
t(3;11)(q21;q12)	3	130,755,505	130,755,497
	11	61,032,619	61,032,619
t(7;13)(q31.3;q21.3)	7	121,245,446	121,438,731
	13	71,023,877	71,023,877

Table 4.1 Comparison of translocation breakpoint positions in the UCSC genome browser between May 2004 and March 2006 builds.

The genome reference sequence from NCBI Build 36 was used for all bioinformatic analysis discussed in this chapter.

4.2 Breakpoints and phenotypes

Translocation breakpoints can affect a patient's phenotype by directly disrupting a gene and affecting it's function, by distancing a gene from its regulatory elements, or placing it under the control of other regulatory elements.

4.2.1 Patient phenotypes (as published in (Gribble et al. 2005))

4.2.1.1 Patient t(2;7)(q37.3;p15.1) phenotype

This patient was referred at the age of eight months with mild generalised developmental delay, a slightly beaked nose, and adducted thumbs. At the age of two years and eight months his development was clearly delayed: he was able to sit but unable to walk unaided and he had no intelligible words. He had a dysmorphic appearance with brachycephaly, blepharophimosis, medially flared eyebrows, a broad nasal tip, short philtrum, thin upper lip, and prominent lower jaw.

4.2.1.2 Patient t(3;11)(q21;q12) phenotype

This patient is one of phenotypically discordant, monochorionic, monoamniotic twins born at 29 weeks' gestation. She had a congenital duodenal obstruction requiring surgery, complex congenital heart disease, and facial dysmorphism. Her twin sister who carries the same apparently balanced t(3;11) is clinically normal.

4.2.1.3 Patient t(7;13)(q31.3;q21.3) phenotype

This patient was referred at the age of six years because of developmental delay and autistic features. By the age of seven years, he had epilepsy, learning difficulties, disordered speech and language, and an autism spectrum disorder.

4.2.2 Direct disruption of a gene by the translocation breakpoint

The underlying genetic cause of some diseases has been found by studying patients with translocations who present with the same characteristics as patients showing deletions or duplications in the same region. For example, the *rnex40* gene identified as a candidate for DiGeorge syndrome by the analysis of a t(2;22)(q14;q11.21) translocation and the *ProSAP2* gene as a candidate for terminal 22q13.3 deletion syndrome by the investigation into a t(12;22)(q24.1;q13.3) translocation (Budarf et al. 1995 Bonaglia et al. 2001).

4.2.2.1 Direct gene disruption in patient t(2;7)(q37.3;p15.1)

Analysis of the two breakpoints in NCBI Build 36 revealed that the chromosome 2 breakpoint directly disrupted Centaurin, Gamma 2 (*CENTG2*), whilst the chromosome 7 breakpoint did not directly disrupt a gene (Figure 4.1).

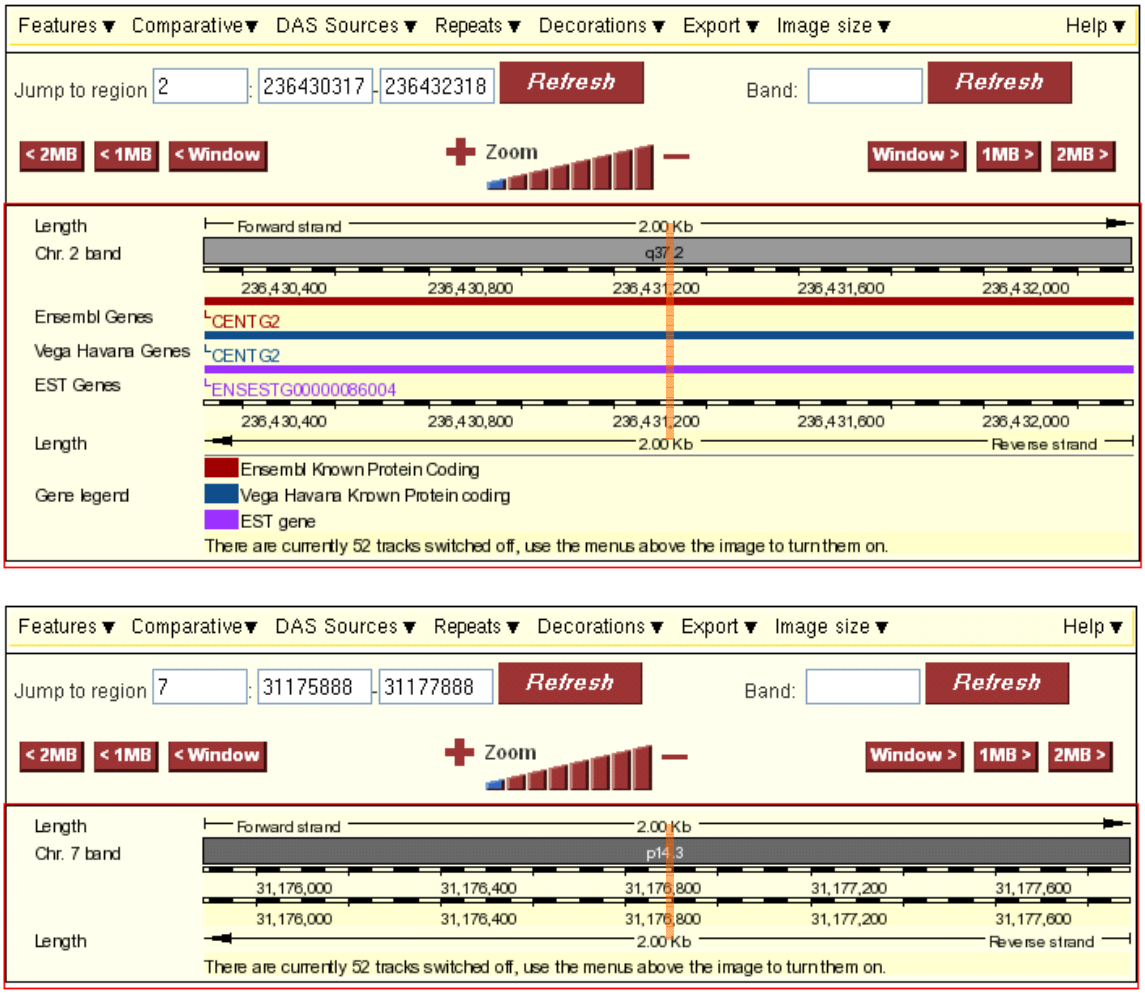


Figure 4.1 NCBI Build 36 Ensembl downloads for the 2Kb region surrounding the chromosome 2 and chromosome 7 breakpoints for patient t(2;7)(q37.3;p15.1). The vertical orange bar marks the breakpoint position.

CENTG2 (OMIM; 608651) is a gene belonging to a protein family involved in membrane trafficking and actin cytoskeleton dynamics – the ADP-ribosylation

GTPase-activating (ARF-GAP) family (Nie et al. 2002; Meurer et al. 2004). *CENTG2* has been considered to be a good candidate for developmental delay as it is expressed in the brain and nervous system, however the evidence for its involvement remains inconclusive (Wassink et al. 2005). Analysis of the chromosome 2 breakpoint in patient t(2;7)(q37.3;p15.1) showed that the breakpoint lay within the intron between exons 9 and 10 of *CENTG2*, resulting in a truncated protein.

4.2.2.2 Direct gene disruption in patient t(3;11)(q21;q12)

Analysis of the breakpoints for patient t(3;11)(q21;q12) showed that neither of the translocation breakpoints directly disrupted a gene (Figure 4.2).

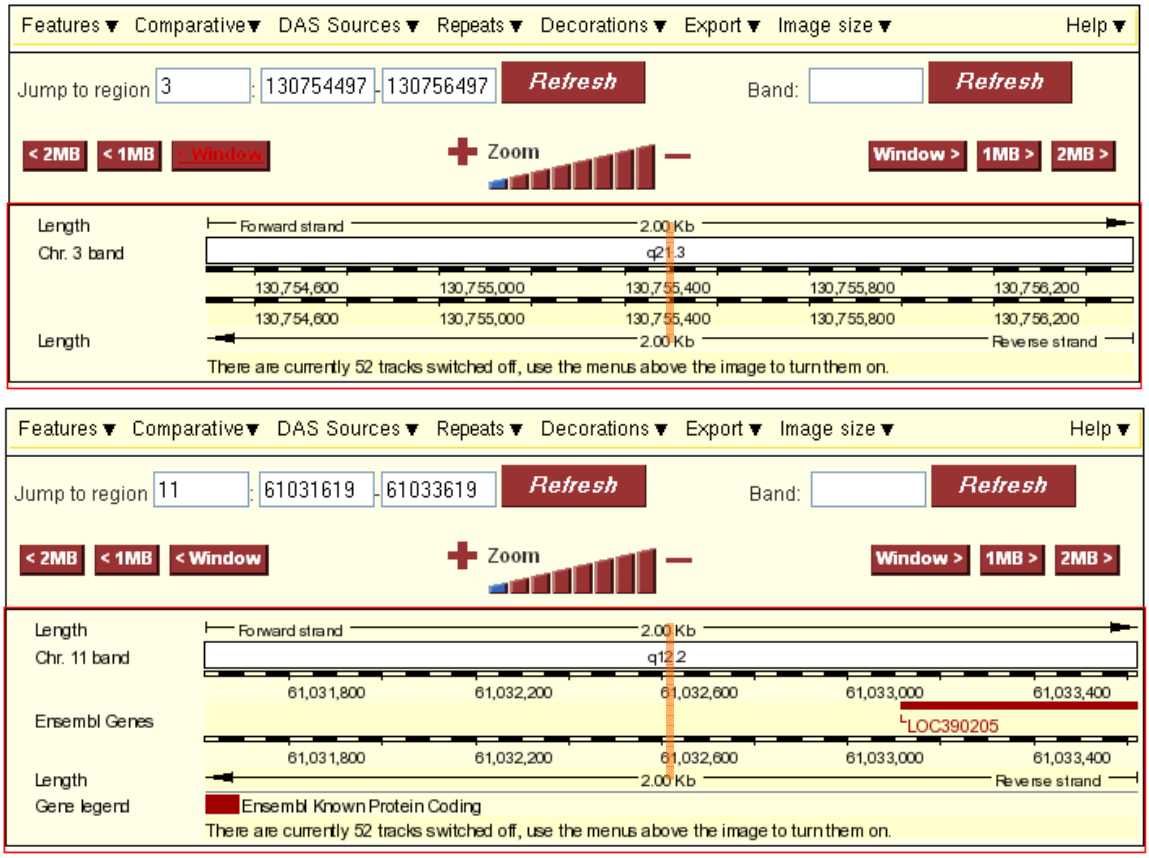


Figure 4.2 NCBI Build 36 Ensembl downloads for the 2Kb region surrounding the chromosome 3 and chromosome 11 breakpoints for patient t(3;11)(q21;q12). The vertical orange bar marks the breakpoint position.

4.2.2.3 Direct gene disruption in patient t(7;13)(q31.3;q21.3)

Analysis of the derivative chromosome sequences showed that the chromosome 7 breakpoint directly disrupted the Protein-Tyrosine Phosphatase, Receptor-Type, Zeta-1 gene (*PTPRZ1*) (Figure 4.3) and the chromosome 13 breakpoint directly disrupted the Dachshund, Drosophila, Homolog of, 1 gene (*DACH1*) (Figure 4.4).

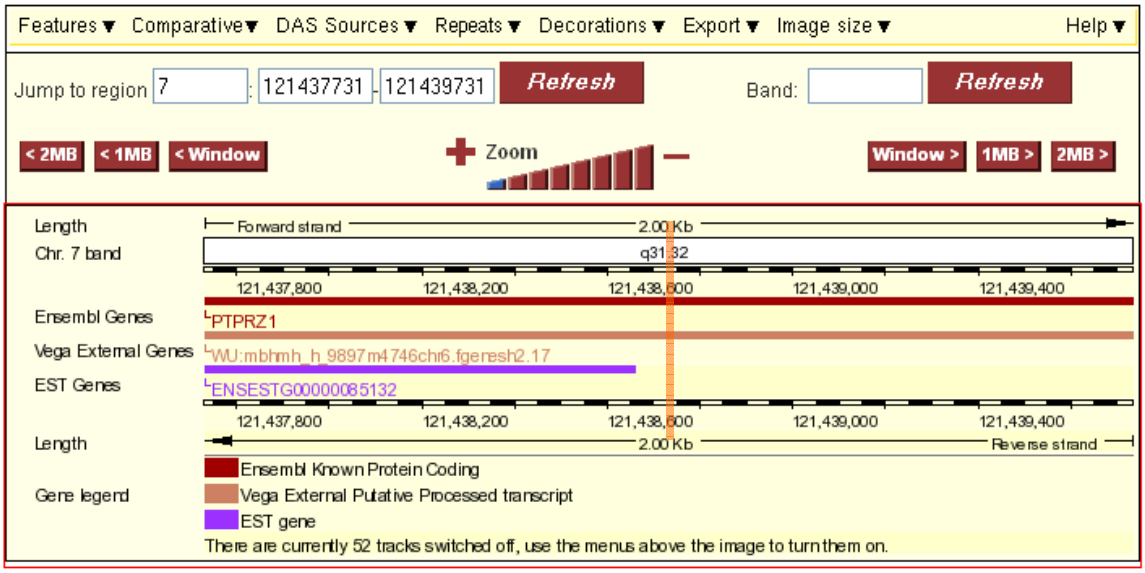


Figure 4.3 NCBI Build 36 Ensembl download for the 2Kb region surrounding the chromosome 7 breakpoint for patient t(7;13)(q31.3;q21.3). The vertical orange bar marks the breakpoint position.

PTPRZ1 (OMIM; 176891) is a member of the Protein Tyrosine Phosphatase (PTP) family and is expressed only in the central nervous system (Levy et al. 1993). There is a high level of expression in the embryonic CNS suggesting that it plays a key role in its development, and it is also expressed in the adult brain at sites of mitotic activity.

PTPRZ1 has been studied in direct relation to its links with autism however the study did not provide any evidence that it is causal to the autism spectrum

disorder (Bonora et al. 2005). This research was supported by a further study in the Japanese population (Marui et al. 2005).

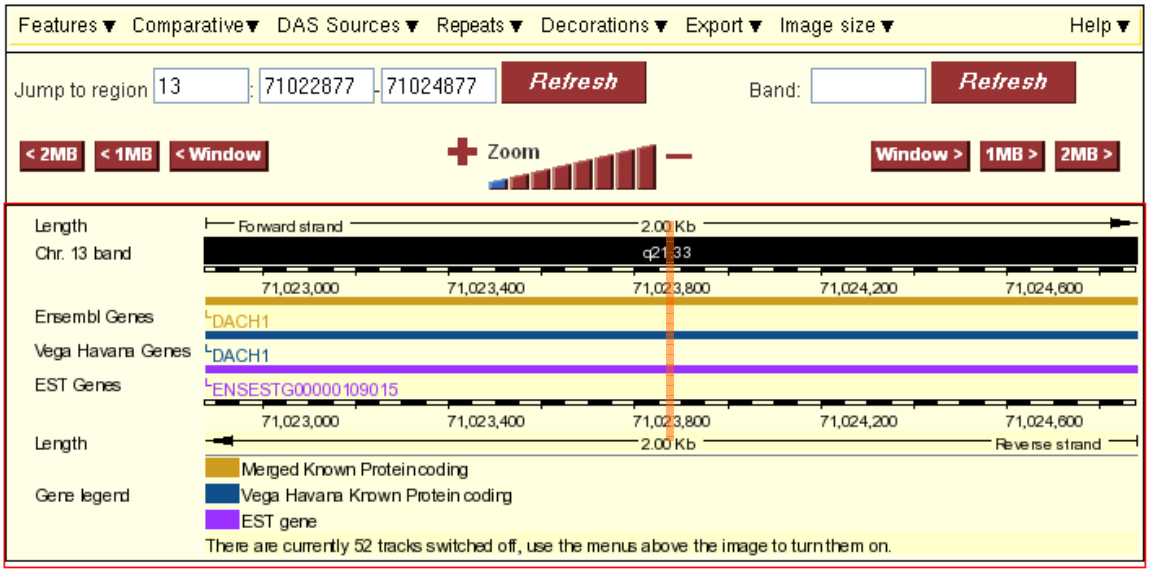


Figure 4.4 NCBI Build 36 Ensembl download for the 2Kb region surrounding the chromosome 13 breakpoint for patient t(7;13)(q31.3;q21.3). The vertical orange bar marks the breakpoint position.

DACH1 (OMIM; 603803) is the human homologue of the Drosophila 'dac' gene involved in leg and eye development (Hammond et al. 1998). As well as playing a key role in development, *DACH1* has been shown to inhibit transforming growth factor-ß (TGF- ß) induced apoptosis (Wu et al. 2003).

Analysis of both breakpoints in relation to the gene structures showed that the chromosome 7 breakpoint disrupted *PTPRZ1* within Exon 12, and the chromosome 13 breakpoint disrupted *DACH1* within Intron 7-8.

4.2.3 Translocation breakpoints and position effect

Positional cloning studies using balanced translocations have shown that in approximately 10% of cases, the breakpoints fall outside of the candidate causative gene (Crisponi et al. 2004). The furthest documented distance of a translocation breakpoint exerting an effect on a gene is 1.3Mb away in a patient with Campomelic Dysplasia (Velagaleti et al. 2005). Tools for the prioritisation of candidate genes in relation to a given phenotype have been developed to aid in genotype to phenotype correlation. Endeavour is a web-based system which has previously identified a novel gene in a 2Mb region involved in craniofacial development which is deleted in some patients with DiGeorge like birth defects (Aerts et al. 2006). The DECIPHER database will also prioritise genes in a region according to published data linking the gene to known characteristics using precise clinical diagnoses.

4.2.3.1 Position effect in patient t(2;7)(q37.3;p15.1)

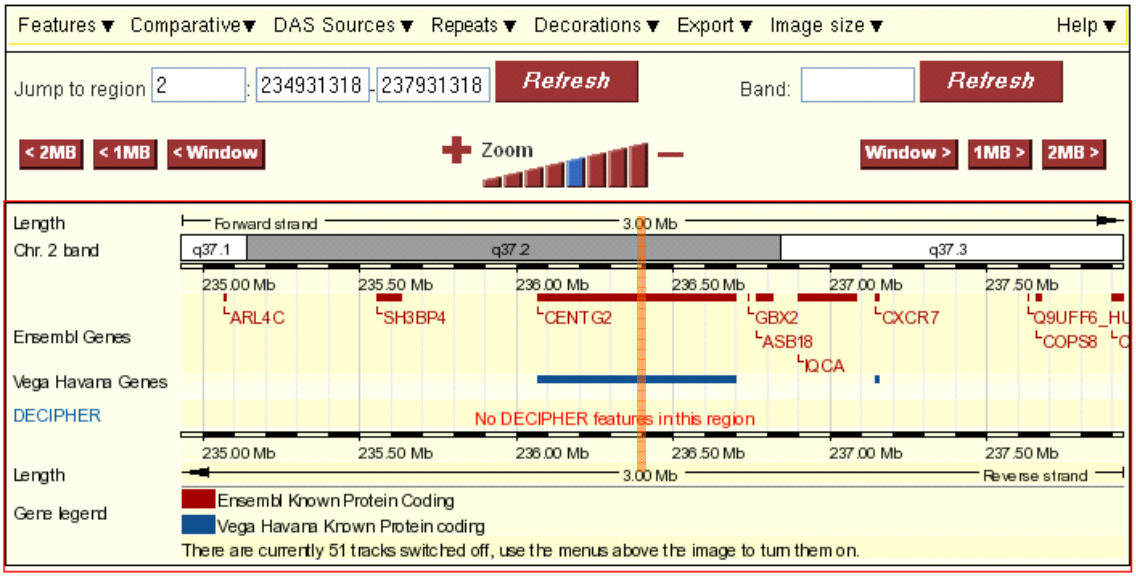


Figure 4.5 NCBI Build 36 Ensembl download for the 3Mb region surrounding the chromosome 2 breakpoint for patient t(2;7)(q37.3;p15.1). The vertical orange bar marks the breakpoint position.

The 3Mb region surrounding the chromosome 2 breakpoint encompasses a total of 10 genes (Figure 4.5). Although Ensembl shows that there are no DECIPHER features there is a common 2q37.3 deletion syndrome in this region. Patients have a range of characteristics, but generally present with varying degrees of facial dysmorphism and mental retardation (Giardino et al. 2001; Aldred et al. 2004; Lukusa et al. 2004) similar to those exhibited by patient t(2;7)(q37.3:p15.1). The exact genetic basis of each characteristic has yet to be determined.

Analysis of the 3Mb region around the chromosome 7 breakpoint identified 24 genes within the potential range for a position effect (Figure 4.6).

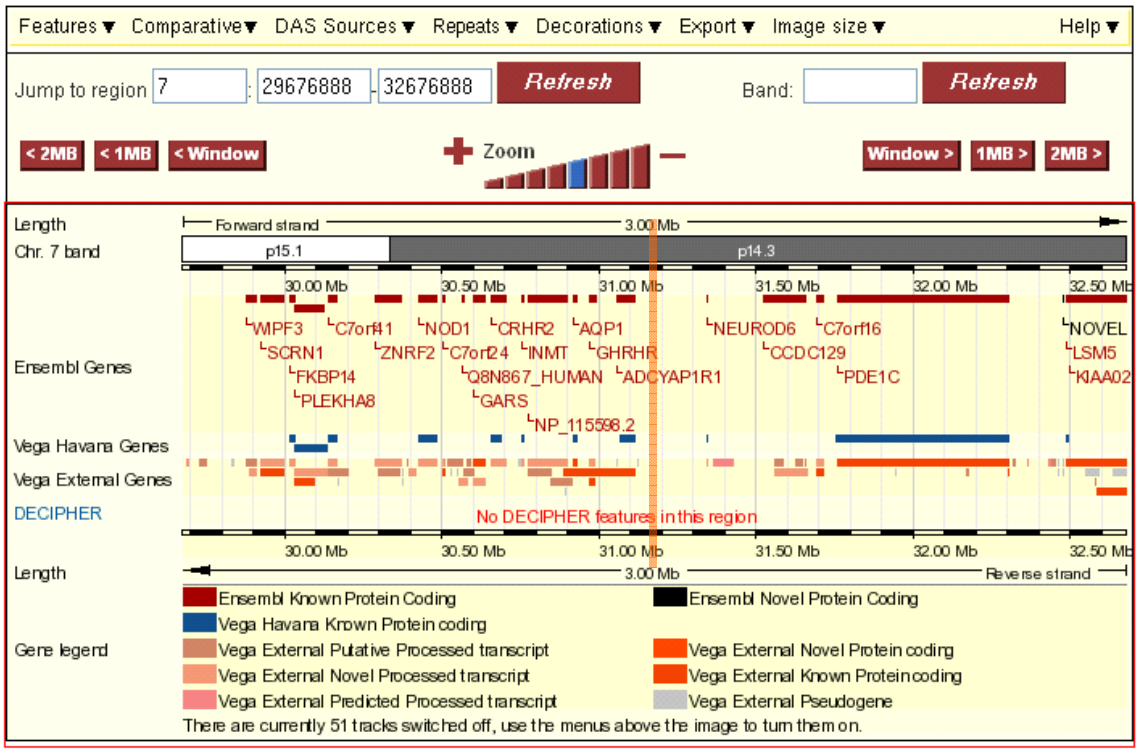


Figure 4.6 NCBI Build 36 Ensembl download for the 3Mb region surrounding the chromosome 7 breakpoint for patient t(2;7)(q37.3;p15.1). The vertical orange bar marks the breakpoint position.

In order to prioritise the genes according to their relevance to the patients phenotype, the DECIPHER database was used. The precise clinical characteristics were entered along with the 3Mb region around the translocation breakpoints for both chromosomes. A strong candidate gene for the patient's phenotype around the chromosome 2 breakpoint is the Collagen Type VI alpha 3 (*COL6A3*) gene. Mutations in *COL6A3* (OMIM; 120250) have been linked to general myopathy of varying severity (Pan et al. 1998; Demir et al. 2002). Analysis of the chromosome 7 breakpoint region using the DECIPHER prioritisation tool identified Aquaporin 1 (*AQP1*) as a candidate gene. *AQP1* (OMIM; 107776) encodes a membrane protein involved in water transport (Smith et al. 1993).

4.2.3.2 Position effect in patient t(3;11)(q21;q12)

The 3Mb region surrounding the chromosome 3 breakpoint contained 43 genes and the chromosome 11 breakpoint, 91 genes. The phenotypes for patient t(3;11)(q21;q12) were ill defined, so no useful data was obtained from the DECIPHER prioritisation tool and no other cases of rearrangement were identified from the database.

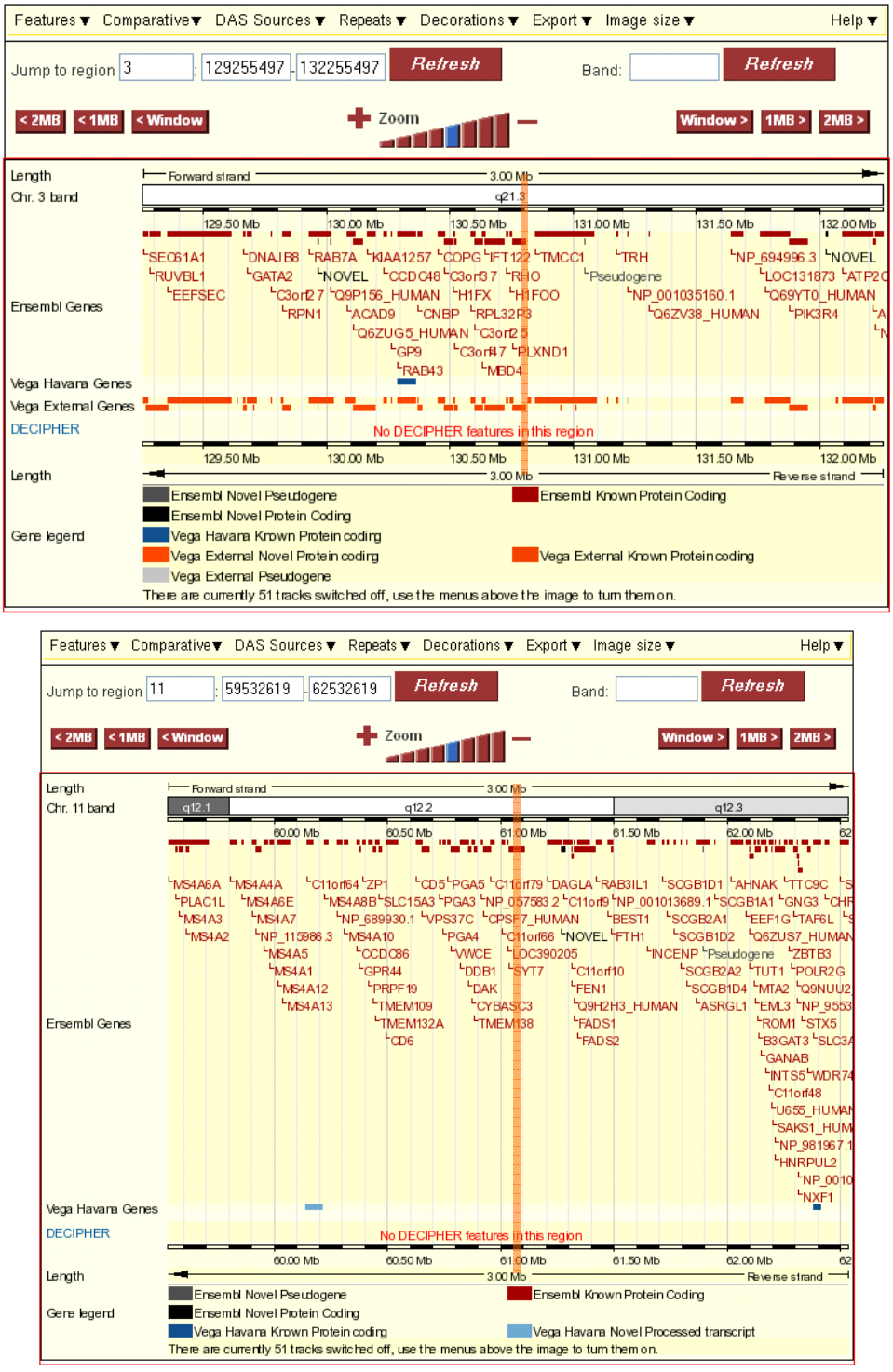


Figure 4.7 NCBI Build 36 Ensembl download for the 3Mb region surrounding A the chromosome 3 breakpoint and B the chromosome 11 breakpoint for patient t(3;11)(q21;q12). The vertical orange bar marks the breakpoint position.

4.2.3.3 *Position effect in patient t(7;13)(q31.3;q21.3)*

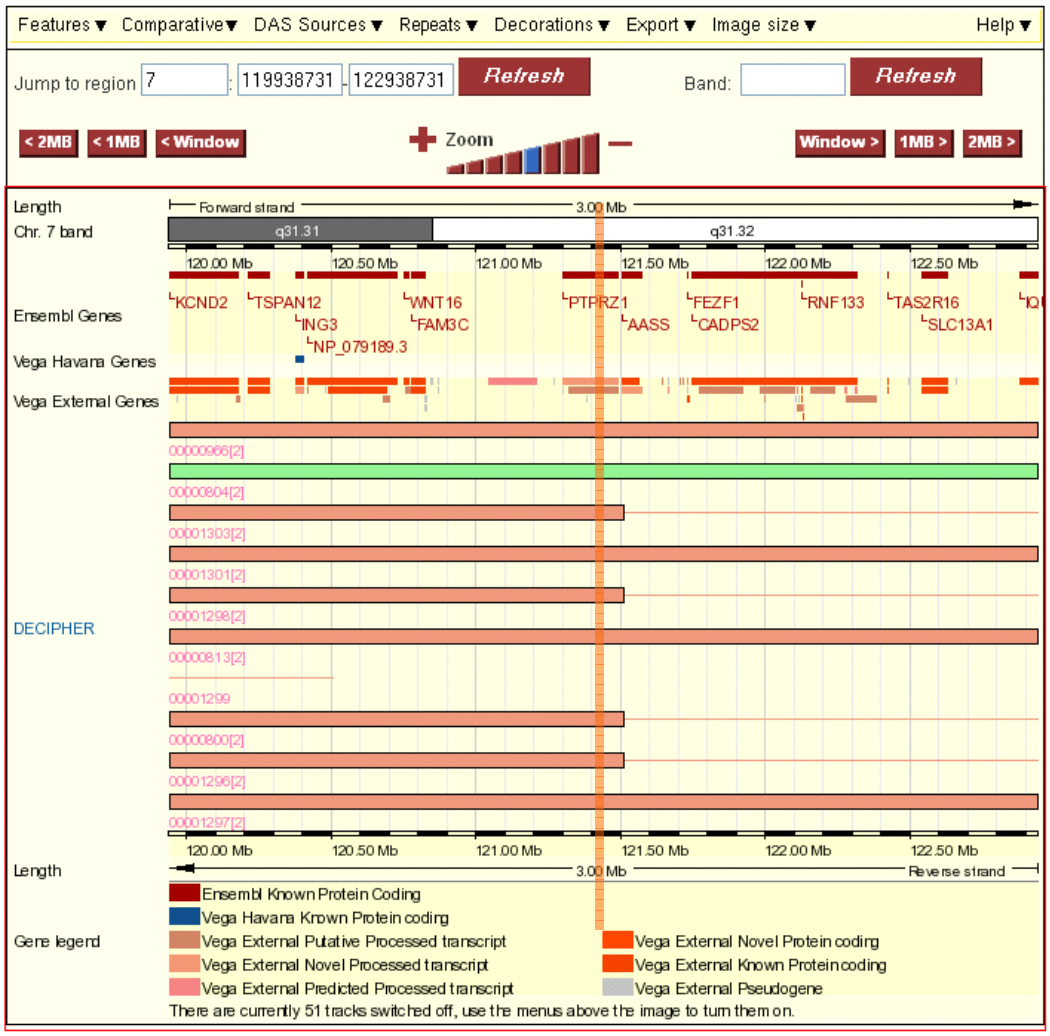


Figure 4.8 NCBI Build 36 Ensembl download for the 3Mb region surrounding the chromosome 7 breakpoint for patient t(7;13)(q31.3;q21.3). The vertical orange bar marks the breakpoint position.

The Ensembl download shows that multiple patients have been entered into the DECIPHER database around the chromosome 7 breakpoint region (Figure 4.8). These patients exhibit characteristics similar to patient t(7;13)(q31.3;q21.3) including developmental delay and autistic features, however so far, no genes have been definitively identified as being causative to the phenotypes.

For the chromosome 13 breakpoint region (Figure 4.9), no patients have been entered in to DECIPHER. The phenotype for this patient has not been defined enough for DECIPHER to prioritise genes, however both the chromosome 7 and 13 breakpoints directly disrupt candidate genes as discussed in Section 4.2.2.3.

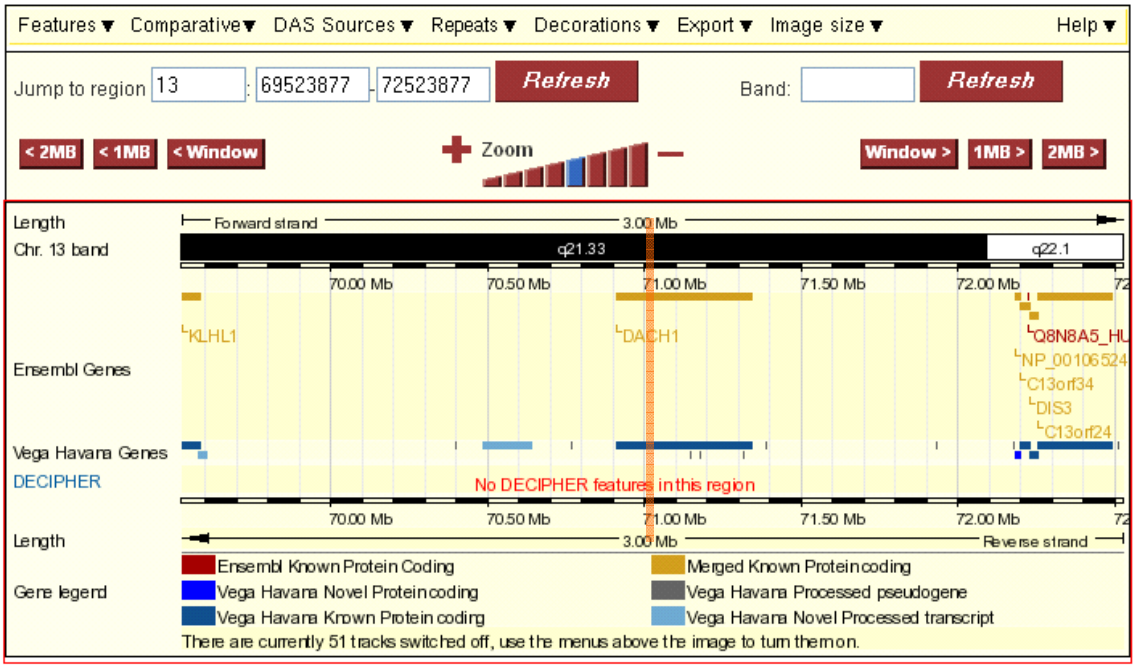


Figure 4.9 NCBI Build 36 Ensembl download for the 3Mb region surrounding the chromosome 13 breakpoint for patient t(7;13)(q31.3;q21.3). The vertical orange bar marks the breakpoint position.

4.2.4 Duplication in patient t(2;7)(q37.3;p15.1) and phenotype

The region of chromosome 3 duplicated in patient t(2;7)(q37.3;p15.1) lies between two Alu repeats at 1.7Mb and 3.6Mb and was found to be approximately 1.9Mb in size (Figure 4.10). Whilst the precise breakpoints could not be determined at the basepair level, the distal breakpoint was mapped to a 260bp region from 1,756,993 to 1,757,252bp and the proximal breakpoint to a 258bp

region from 3,614,132 to 3,614,389bp as discussed in Chapter 3 (there is no variation in the genomic positions for chromosome 3 between NCBI Builds 35 and 36).

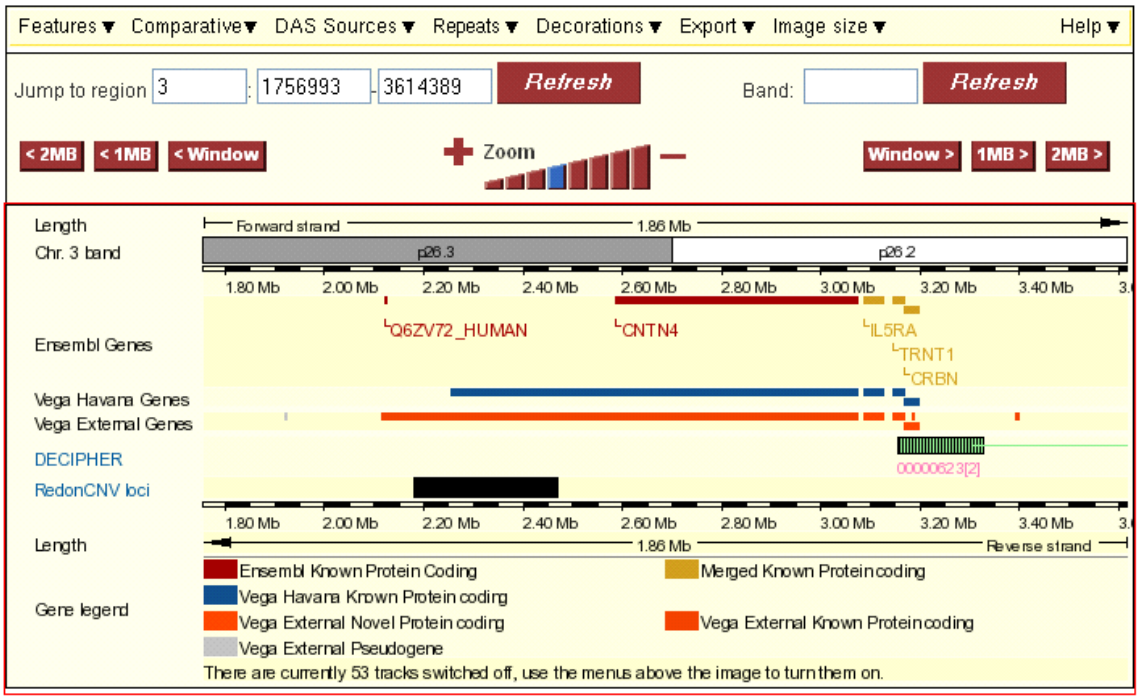


Figure 4.10 Ensembl download showing the duplicated region at 3p26.2-3p26.3 for patient t(2;7)(q37.3;p15.1).

This region contains a known CNV locus at approximately 2.3Mb along the chromosome. 2 out of 270 phenotypically normal individuals used to generate this CNV data showed amplification at this locus. A literature search revealed reports of several aberrations in this region of the genome; the most common being a 3p deletion syndrome characterised by developmental delay, growth retardation and dysmorphic features (Neri et al. 1984; Moncla et al. 1995; Lukusa et al. 1999; Wahlstrom et al. 1999; Cargile et al. 2002; Fernandez et al. 2004; Lalli et al. 2007). The DECIPHER track for this region reports a patient with a 1.4Mb duplication of dup(3)(p26.1;p26.2) presenting with an absent uterus and

developmental delay. There was only one published report of a duplication case covering the same region; the patient showed a two clone duplication (by array CGH on a 1Mb resolution microarray) and presented with dysmorphic features and learning disabilities (Shaw-Smith et al. 2004). However, as was the case for patient t(2;7)(q37.3;p151), the duplication was inherited from a phenotypically normal father. It is unlikely that the duplication plays a major part in the proband's phenotype – it is more likely that the translocation breakpoints are responsible for the phenotype as discussed in 4.2.2.1 and 4.2.3.1.

4.3 Breakpoints and genome architecture

Investigations into recurrent rearrangements have shown that genomic structures may pre-dispose the sequence to rearrangement. For example; Sotos syndrome is mediated by low copy repeats (LCRs) (Kurotaki et al. 2005; Visser et al. 2005), the duplications associated with Charcot-Marie-Tooth disease type 1A and the deletions associated with hereditary neuropathy with liability to pressure palsies are both mediated by LCRs on chromosome 17p12 (Shaw et al. 2004) and the recurrent t(11;22) translocation is mediated by AT-rich palindromic repeats (Edelmann et al. 2001; Kurahashi and Emanuel 2001). However, the mechanisms underlying *de novo* constitutional translocations have yet to be elucidated.

4.3.1 Analysis of sequence at breakpoints

The sequence data available for constitutional non-recurrent reciprocal translocation breakpoints in the literature is summarised in Table 4.2. The majority of these translocations are accompanied by the loss, addition or duplication of a small number of bases, believed to be a characteristic of non-homologous end joining as discussed in the Introduction.

Translocation	insertions at bpt	deletions at bpt	duplications at bpt	Reference
t(X;21)(p21;p12)	3bp on derX	71-72bp from chrX 16-23bp from chr21		Bodrug et al., 1987
t(X;2)(p21;q37)	2bp on der2	1-3bp from chr2 0-2bp from chrX		Bodrug et al., 1991
t(X;4)(p21;q35)	3bp on derX	2-3bp from chrX 7-8bp from chr4		
t(X;1)(p21;p34)	2-5bp on derX	4-7bp from chrX		Cockburn, 1991
t(X;4)(p21.2;q31.22)		5Kb from chrX 4-6bp from chr4		Giacalone and Francke, 1992
t(4;22)(q12;q12.2)				Arai, Ikeuchi and Nakamura, 1994
t(2;22)(q14;q11.21)		5-10bp from chr2 1-6bp from chr22		Budarf et al., 1995
t(X;5)(p21;q31.1)	4bp on derX 6bp on der5	3bp from chrX		van Bakel et al., 1995
t(X;9)(p21.1;q34.3)	7bp on derX 3bp on der9	1bp from chrX 55bp from chr9		Toriello et al., 1996
t(21;22)(p12;q11)	1bp on der21 or der22			Holmes et al., 1997
t(X;8)(p22.13;q22.1)				Ishikawa_Brush et al., 1997
t(17;22)(q11.2;q11.2)	109bp on der17	14bp from chr17		Kehrer-Sawatzki et al., 1997
t(6;7)(q16.2;p15.3)		518bp from chr7		Krebs et al., 1997
t(8;17)(p11.2;p13.3)			3bp of chr8	Kurahashi et al., 1998
t(1;10)(p22;q21)				Roberts, Chernova and Cowell, 1998
t(2;19)(q11.2;q13.3)		1bp from chr2 2bp from chr19		Yoshiura et al., 1998
t(6;12)(q16.2;q21.2)	15bp on der12	6bp from chr12	5bp of chr6	Ikegawa et al., 1999
t(1;6)(p22.1;q16.2)		1bp from chr6		Holder, Butte and Zinn, 2000
t(1;8)(q21.1;q22.1)			6bp of chr8	Matsumoto et al., 2000
t(1;11)q42.1;q14.3)	2bp on der11	4bp from chr11		Millar et al., 2000
t(12;22)(q24.1;q13.3)			5bp of chr22	Bonaglia et al., 2001
t(1;19)(q21.3;q13.2)				Nothwang et al., 2001
t(9;11)(p24;q23)	41bp on der9	2bp from chr11		Willett-Brozick et al., 2001
t(7;16)(q11.23;q13)				Duba et al., 2002
t(1;8)(p34.3;q21.12)	5bp on der1 12bp on der8	1bp from chr1 10bp from chr8		McMullan et al., 2002
t(2;8)(q31;p21)	5bp on der2 13bp on der8	7bp from chr8		Spitz et al., 2002
t(2;8)(q31;p21)	13bp on der8	2bp from chr8		Sugawara et al., 2002
t(6;13)(q21;q12)			2bp of chr6	Vervoort et al., 2002
t(7;22)(p13;q11.2)	13bp on der7 46bp on der22	75bp from chr7 4bp from chr22		Hill et al., 2003
t(6;11)(q14.2;q25)		8bp from chr6 9bp from chr11	16bp of chr11	Jeffries et al., 2003
t(4;22)(q35.1;q11.2)		1bp from chr4 ~168bp from chr22		Nimmakayalu et al., 2003
t(X;7)(p11.3;q11.21)				Shoichet et al., 2003
t(1;7)(q41;p21)		3-6bp from chr1 3-6bp from chr7		David et al., 2003
t(1;22)(p21.2;q11)				Gotter et al., 2004
t(3;8)(p14.2;q24.2)		5Kb from chr3		Rodriguez-Perales et al., 2004
t(2;6)(q24.3;q22.31)			18bp of chr6	Bocciardi et al., 2005
t(4;17)(q28.3;q24.3)		10-13bp from chr4 0-3bp from chr17		Velagaleti et al., 2005
t(1;7)(p22;q32)	2bp on der1 3bp on der7			Borg et al., 2005
t(4;15)(q27;q11.2)	1bp on der4			Schule et al., 2005
t(4;15)(q22.3;q21.3)	7bp on der4 37bp on der15	1bp from chr4 13bp from chr15		Klar et al., 2005
t(9;11)(q33.1;p15.3)				Tagariello et al., 2006
t(6;17)(p21.31;q11.2)		2bp from chr6	7bp of chr17	Mansouri et al., 2006
t(5;14)(q21;q32)	2bp on der14	1bp from chr5 1bp from chr14		Haider et al., 2006
t(17;22)(q21.1;q12.1)	4bp on der17			Gribble et al., 2007
t(2;7)(q37.1;q36.3)		11bp from chr2 1bp from chr7		
t(11;17)(p13;p13.1)	3bp on der11	2bp from chr11		
t(2;7)(q37.1;q21.3)				Bocciardi et al., 2007

Table 4.2 Summary of sequence data at the breakpoints of published reciprocal constitutional translocations.

4.3.1.1 Patient t(2;7)(q37.3;p15.1)

Alignment of the sequence across the two breakpoint regions revealed a 7bp deletion of chromosome 2 material, a 4bp duplication of chromosome 7 sequence and a 19bp insertion of unknown origin at the derivative 7 junction (Figure 4.11). An example of mitochondrial DNA insertion at the breakpoints has been observed in a familial t(9;11)(p24;q23) translocation (Willett-Brozick et al. 2001) however this is not the origin of the 19bp insertion for patient t(2;7)(q37.3;p15.1).

```
t(2;7)(q37.3;p15.1)
2r/c;  aaatgaaaatcccccaagatcacccaggaagcagccctgggtataagcctcacactcccaacat
chr2;  actgtctgcccatgttgggagtgtgaggcttatacccagggctgcttcctgggtgatcttgggg
der2;  actgtctgcccatgttgggagtgtgaggcttataTTAAGGCTTTAAATATCTGAATGTAGCCAT
der7;  cagcccaggaagcagtttaatctattaatctttgTTAATAGATTAAACAAAAACTGAGCGTAAA
chr7;  GGTACCTCATGGCTACATTCAGATATTTAAAGCCTTAATAGATTAAACAAAAACTGAGCGTAAA
7r/c;  ACCCTGAGTTTACGCTCAGTTTTTGTTTAATCTATTAAGGCTTTAAATATCTGAATGTAGCCAT
```

Figure 4.11 Sequence alignment of translocation junctions against human reference sequence for patient t(2;7)(q37.3;p15.1).

4.3.1.2 Patient t(3;11)(q21;q12)

In this patient there is a 1-3bp deletion of chromosome 3 sequence, and a 8-10bp deletion of chromosome 11 sequence (Figure 4.12). The precise number of basepairs involved cannot be determined due to the 2 bp homology (GC) at the breakpoints of both chromosomes.

```
t(3;11)(q21;q12)
chr3   acaccctgatcctgagttcactcctcggcccgcccatccaagcgccccgtgcggctggcgtt
der3;  acaccctgatcctgagttcactcctcggcccGGGCTGGATGGGCAGGTAGGGGGCGGGCTCCGG
der11; CGGGGCGGGGAATCTCTCGGCTTGTGCTTGCcccatccaagcgccccgtgcggctggcgttctc
chr11; CGGGGCGGGGAATCTCTCGGCTTGTGCTTGCTCCGCGGTGGGCTGGATGGGCAGGTAGGGGGCG
```

Figure 4.12 Sequence alignment of translocation junctions against human reference sequence for patient t(3;11)(q21;q12).

4.3.1.3 Patient t(7;13)(q31.3;q21.3)

In this patient there is a 2-3bp deletion of chromosome 7 sequence and a 4-5bp duplication of chromosome 13 sequence (Figure 4.13). The exact number cannot be determined due to the T at the breakpoints being present on both chromosomes.

```
t(7;13)(q31.3;q21.3)
chr7;  tagtgattcggccttgcatgctacgcctgtattcccagtgtcgatgtgtcatttgaatccatc
der7;  tagtgattcggccttgcatgctacgcctgGCTTTGTATGAAAATGGTCCATAAGATTGGATGCT
der13; GTGATACATAATGATTTTCTTTCAAAAGGGCTTTttcccagtgtcgatgtgtcatttgaatcca
chr13; GTGATACATAATGATTTTCTTTCAAAAGGGCTTTGTATGAAAATGGTCCATAAGATTGGATGCT
```

Figure 4.13 Sequence alignment of translocation junctions against human reference sequence for patient t(7;13)(q31.3;q21.3).

4.3.1.4 Summary of basepair sequence analysis at the breakpoints

Analysis of the derivative chromosome sequence for patients t(2;7)(q37.3;p15.1), t(3;11)(q21;q12) and t(7;13)(q31.3;q21.3) revealed that a small number of bases were deleted (1-10bp), duplicated (4-5bp) and/or inserted (19bp) in each case. This is comparable to the numbers seen in the published cases where deletions of 1-518bp, duplications of 1-18bp and insertions of 2-109bp were observed. These small basepair changes are indicative of non-homologous end joining as a mechanism for the rearrangements.

4.3.2 Analysis of repeat structures and recombination motifs at breakpoints

As described in the Introduction, certain repeat structures and recombination motifs have been associated with chromosomal rearrangements. Recurrent chromosomal rearrangements are frequently associated with repeat structures within the genome, whilst the mechanisms behind non-recurrent rearrangements remain undetermined. Analysis of the sequence surrounding the translocation breakpoints in the 3 patients studied in this thesis, and a comparison with non-recurrent translocations in the literature may help to elucidate a mechanism.

Repeat structures present at or around the breakpoints of published non-recurrent constitutional translocations are summarised in Table 4.3.

Translocation	Repeats in proximity	Reference
t(X;21)(p21;p12)		Bodrug et al., 1987
t(X;2)(p21;q37)	ALU 50bp from chr 2bp	Bodrug et al., 1991
t(X;4)(p21;q35)		
t(X;1)(p21;p34)		Cockburn, 1991
t(X;4)(p21.2;q31.22)		Giacalone and Francke, 1992
t(4;22)(q12;q12.2)		Arai, Ikeuchi and Nakamura, 1994
t(2;22)(q14;q11.21)	Chi like octamer close to chr22 bpt	Budarf et al., 1995
t(X;5)(p21;q31.1)		van Bakel et al., 1995
t(X;9)(p21.1;q34.3)	LINE elements on chrX	Toriello et al., 1996
t(21;22)(p12;q11)		Holmes et al., 1997
t(X;8)(p22.13;q22.1)		Ishikawa_Brush et al., 1997
t(17;22)(q11.2;q11.2)	AT rich repeats on both chrs	Kehrer-Sawatzki et al., 1997
t(6;7)(q16.2;p15.3)	chr7 breakpoint within ALU repeat	Krebs et al., 1997
t(8;17)(p11.2;p13.3)	5xALU repeats distal to chr17 breakpoint 3xL1distal to chr8 breakpoint	Kurahashi et al., 1998
t(1;10)(p22;q21)		Roberts, Chernova and Cowell, 1998
t(2;19)(q11.2;q13.3)	ALU repeat on chr19	Yoshiura et al., 1998
t(6;12)(q16.2;q21.2)		Ikegawa et al., 1999
t(1;6)(p22.1;q16.2)		Holder, Butte and Zinn, 2000
t(1;8)(q21.1;q22.1)		Matsumoto et al., 2000
t(1;11)q42.1;q14.3)		Millar et al., 2000
t(12;22)(q24.1;q13.3)		Bonaglia et al., 2001
t(1;19)(q21.3;q13.2)	chr1 bpt within AluSp chr19 bpt within AluY	Nothwang et al., 2001
t(9;11)(p24;q23)	chr9 bpt within L1 repeat chr11 bpt within ALU repeat	Willett-Brozick et al., 2001
t(7;16)(q11.23;q13)		Duba et al., 2002
t(1;8)(p34.3;q21.12)		McMullan et al., 2002
t(2;8)(q31;p21)		Spitz et al., 2002
t(2;8)(q31;p21)		Sugawara et al., 2002
t(6;13)(q21;q12)		Vervoort et al., 2002
t(7;22)(p13;q11.2)	chr22 bpt within Immunogolbulin Lambda light chain locus	Hill et al., 2003
t(6;11)(q14.2;q25)	chr6 bpt within LINE L1 element	Jeffries et al., 2003
t(4;22)(q35.1;q11.2)	554bp palindrome on chr 4 PATTR on chr 22	Nimmakayalu et al., 2003
t(X;7)(p11.3;q11.21)	LINE repeats at both bpts (L1ME on chrX, L1 on chr7)	Shoichet et al., 2003
t(1;7)(q41;p21)	chr1 bpt 200bp proximal to L1 element chr7 bpt within L2 element	David et al., 2003
t(1;22)(p21.2;q11)	AT rich region on chr1 LCR22 on chr22 Palindromic repeats on both chrs	Gotter et al., 2004
t(3;8)(p14.2;q24.2)	chr3 bpt in AT rich region	Rodriguez-Perales et al., 2004
t(2;6)(q24.3;q22.31)	SINE and LINE elements lie close to both breakpoints	Bocciardi et al., 2005
t(4;17)(q28.3;q24.3)	chr4 bpt within mariner-transposon like element (HSMAR2)	Velagaleti et al., 2005
t(1;7)(p22;q32)		Borg et al., 2005
t(4;15)(q27;q11.2)	chr4 bpt within LTR1B chr15 bpt surrounded by SINE, AluY and LINE;L1M4	Schule et al., 2005
t(4;15)(q22.3;q21.3)	polypurine and polypyrimidine tracts around both breakpoints	Klar et al., 2005
t(9;11)(q33.1;p15.3)		Tagariello et al., 2006
t(6;17)(p21.31;q11.2)	both breakpoints lie within polypyrimidine tracts	Mansouri et al., 2006
t(5;14)(q21;q32)		Haider et al., 2006
t(17;22)(q21.1;q12.1)	chr22 bpt within SINE (MIR)	Gribble et al., 2007
t(2;7)(q37.1;q36.3)	chr2 bpt 2bp away from LINE (L2) chr7 bpt within LINE (L1)	
t(11;17)(p13;p13.1)	both bpts within LINE (L1) repeats	
t(2;7)(q37.1;q21.3)		Bocciardi et al., 2007

Table 4.3 Summary of sequence motifs found at or around the breakpoints of published reciprocal constitutional translocations.

A selection of repeat structures were searched for in the 2Kb regions around the translocation breakpoints in the patients with translocations t(2;7)(q37.3;p15.1), t(3;11)(q21;q12) and t(7;13)(q31.3;q21.3) which were mapped and sequenced in Chapter 3. These structures have been discussed in detail in the Introduction, but briefly include recombination motifs such as Chi sequences, Topoisomerase I and II sites, repetitive DNA sequences such as mini satellite sequences, purine/pyrimidine tracts and AT rich sites, SINE elements including Alu repeats and MIR repeats, LINE elements, long terminal repeats, segmental duplications and DNA motifs such as palindromic sequences. 2Kb regions surrounding the translocation breakpoints for each chromosome were exported from Ensembl (http://www.ensembl.org/Homo_sapiens/index.html) and analysed for these structures. A full list of websites used to investigate the presence of sequence motifs is summarised in Appendix A1. Any structures found to be present are detailed in Sections 4.3.2.1 to 4.3.2.3.

4.3.2.1 Analysis of repeat structures around t(2;7)(q37.3;p15.1) breakpoints

A MER1B repeat was found 526bp proximal to the chromosome 2 breakpoint. The chromosome 7 breakpoint was found to be more repetitive. Two SINE elements were found; an AluSx element 375bp distal and a MIR repeat 526bp proximal to the breakpoint. A total of 4 LINE elements were found; 3 L3 elements 2bp distal and 114bp and 302bp proximal and an L2 element 8bp proximal to the breakpoint. Also found were 6.5 copies of a simple 4bp tandem repeat (AAAG) 375 bp proximal to the breakpoint. A Chi motif was observed 895 bp proximal to the chromosome 2 breakpoint with a Translin motif (ATGCAG) observed 570bp distal to the breakpoint. Translin motifs (ATGCAG) were discovered 546bp and 894 bp proximal to the chromosome 7 breakpoint. DNA bind motifs (TTTAAA) were observed 381bp proximal and 594bp distal to the chromosome 2 breakpoint and 826 and 8bp distal to the chromosome 7 breakpoint.

4.3.2.2 Analysis of repeat structures around t(3;11)(q21;q12) breakpoints

Two LINE elements were found proximal to the chromosome 3 breakpoint; an L2 repeat 900bp from the breakpoint and an L1ME4a repeat 20bp from the breakpoint. In addition, a MIRb SINE element was found 776bp distal to the breakpoint. The chromosome 11 breakpoint was found to lie distal to 2 SINE elements; a MIRb 906bp proximal and a MIR 614bp proximal to the breakpoint. In addition, 128bp of a simple repeat (CGG) was found 292bp distal to the chromosome 11 breakpoint with a further 3.8 copies of a 5bp (GCCCC) repeat and 2 copies of a 9bp repeat (GAGCTGCGC) both distal to the breakpoint by 74 and 621bp respectively. The chromosome 3 sequence was noted to contain a chi motif 279bp proximal, a translin motif (ATGCAG) 544bp distal and 2 translin motifs (GCCC[A/T][G/C][G/C][A/T]) 525bp and 945bp distal to the breakpoint. The chromosome 11 breakpoint was 880bp distal, 563bp distal and 512bp distal to 3 Chi motifs and 616bp distal, 606bp distal, 554bp distal and 794bp proximal to 4 Translin motifs (GCCC[A/T][G/C][G/C][A/T]). The chromosome 11 sequence was noted to contain palindromic sequence from 61,032,591 to 61,032,600bp with 61,032,698 to 61,032,707bp or 61,032,693 to 61,032,702bp. Both of these regions of palindromic sequence surround the breakpoint on chromosome 11. DNA bind motifs were observed 292 and 368bp distal to the chromosome 3 breakpoint and 697bp proximal to the chromosome 11 breakpoint

4.3.2.3 Analysis of repeat structures around t(7;13)(q31.3;q21.3) breakpoints

The chromosome 13 breakpoint was found to lie 130bp distal to an L1MA4 LINE repeat, 46bp distal to a 28bp AT rich low complexity repeat and 154bp proximal to a simple 16bp tandem repeat present in 2 copies. The chromosome 7 breakpoint was 114bp and 550bp proximal to 2 Translin motifs (ATGCAG). In addition, 3 immunoglobulin heptamers were observed; GATAGTG 610bp distal and CACAGTC 810bp distal and 825bp proximal to the chromosome 7 breakpoint. However, the immunoglobulin nonamer motif was not observed in close proximity to the heptamer motifs. The chromosome 13 breakpoint was

128bp proximal to a single Translin motif (ATGCAG). DNA bend motifs were discovered 898bp and 784bp proximal to the chromosome 7 breakpoint and 733, 609, 500 and 94bp proximal and 958bp distal to the chromosome 13 breakpoint.

4.3.2.4 Summary of sequence motifs found at translocation breakpoints

Whilst recombinogenic motifs were found within 1Kb of the breakpoint in all 6 regions studied, none were found to cross the exact breakpoint locations. In addition, repeat structures were observed within 1Kb of the breakpoints in all 6 regions studied, however the same repeat structures were not observed on both donor chromosomes for each patient. Whilst a comparison of the data generated from the 3 patients studied did not appear to implicate any particular repeat structure as being definitively involved in the formation of non-recurrent translocations. However, a study of 75 breakpoint sequences and 5000 control sequences has revealed a significant enrichment of MIR repeats (Kalaitzopoulos 2006).

4.4 Mechanisms underlying genomic rearrangements

Whilst recurrent translocations such as the t(11;22) are believed to be caused by illegitimate homologous recombination, the mechanism behind non-recurrent constitutional translocations has yet to be defined.

4.4.1 Translocations in patients t(2;7)(q37.3;p15.1), t(3;11)(q21;q12) and t(7;13)(q31.3;q21.3)

Analysis of the repeat structures and recombination sites at or around the breakpoints in the 3 patients and their comparison to the structures observed at the breakpoints in published translocation cases did not highlight any structure involved in all cases (as discussed in section 4.3.2.4). No regions of homology were observed at the sites of translocation between the donor chromosomes. In addition, a few bases (1-19bp) were seen to be deleted, duplicated or inserted in the derivative chromosomes of the 3 patients which were comparable to the 1-

518bp observed in the literature. These observations have led to the conclusion that the reciprocal translocations observed in the 3 patients have arisen due to non-homologous end joining.

4.4.2 Duplication in patient t(2;7)(q37.3;p15.1)

Sequence analysis of the dup(3)(p26.3p26.3) breakpoints showed that both breakpoints fell within Alu repeats at 1.7Mb and 3.6Mb along chromosome 3. Due to the high degree of homology (97%) within the Alu elements the distal breakpoint was only mapped within a 260bp region from 1,756,993 to 1,757,252bp and the proximal breakpoint was mapped within a 258bp region from 3,614,132 to 3,614,389bp. The homology observed between both breakpoint regions is indicative of a rearrangement mediated by non-allelic homologous recombination (Figure 4.14).

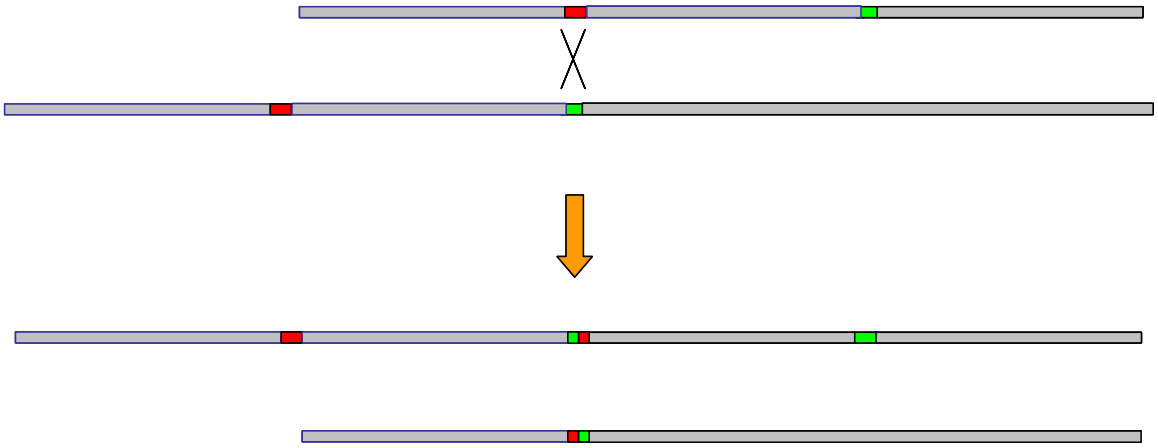


Figure 4.14 Schematic of the non-allelic homologous recombination mechanism proposed for the dup(3)(p26.3p26.3) observed in patient t(2;7)(q37.3;p15.1). Alignment of homologous segments of DNA (in this case direct Alu repeats with 97% homology coloured red and green) can lead to unequal crossing over resulting in a duplication of material on one chromosome and the reciprocal deletion on the other.

4.5 Conclusions

Analysis of the sequence across the translocation breakpoints in all 3 patients revealed that 3 of the 6 breakpoints directly disrupted a gene; CENTG2 on chromosome 2 for patient t(2;7)(q37.3;p15.1) and PTPRZ1 on chromosome 7 and DACH1 on chromosome 13 for patient t(7;13)(q31.1;q21.3), providing candidate genes for the 2 of the 3 patient phenotypes. In addition, analysis of the genes in the surrounding areas susceptible to position effect identified candidate genes for the patient analysed.

Investigation into the genomic architecture revealed that the dup(3)(p26.3p26.3) in patient t(2;7)(q37.3;p15.1) was likely to have resulted from non allelic homologous recombination mediated by 2 Alu repeats with 97% homology. This was in direct contrast to the proposed mechanism of non homologous end joining for the translocations where no repeat structures/recombination motifs were found to be common to all the breakpoints.

5 Results: Investigation of chromosome translocations using custom-made libraries

5.1 Introduction

The chromosome breakpoints of apparently balanced translocations can exhibit additional complexity when studied in detail, in some cases leading to difficulties in generating sequence across the translocation junctions (Puissant et al. 1988; Giacalone and Francke 1992; Borg et al. 2002; Cox et al. 2003; Rodriguez-Perales et al. 2004). Translocation breakpoint mapping by array painting onto microarrays of varying resolution (1Mb, WGTP, fosmid, PCR product, oligonucleotide) followed by long range PCR as described in Chapter 3 has successfully identified the breakpoints for the cases studied in this thesis as these patients carried balanced reciprocal translocations. For cases that are not balanced around the translocation breakpoints, additional methodologies of breakpoint mapping are required. For this reason, we have investigated the generation and screening of custom-made libraries to isolate clones containing the translocation junctions. The sequencing of these clones and the analysis of the sequence generated (approximately 36Kb) will reveal the translocation junction in addition to providing an insight into the complexity of the rearrangement. This chapter assesses 2 methods for creating a custom-made library; one with cosmids (using an in-house protocol) and one with fosmids (using a commercial kit) and the subsequent screening of the library using Alu PCR, PCR product pools and end sequencing.

5.2 Comparison of custom-made library protocols

Two methods for creating custom-made libraries were assessed; the first used enzymes to fragment the DNA which was subsequently ligated into a cosmid vector, and the second used a syringe needle to shear the DNA which was subsequently ligated into a fosmid vector. Both systems have a similar resolution

of approximately 40Kb as they use lambda phage packaging systems. Whole genomic DNA from a patient with a constitutional translocation t(7;13)(q31.3;q21.3) previously characterised by array painting and PCR (Chapter 3) was used to assess these procedures as viable methods of custom-made library production.

5.2.1 Cosmid library

Seven tubes each containing 150ng of DNA were competitively digested using the restriction enzymes *Mbo1* and dam methylase. Different proportions of *Mbo1*: dam methylase were used to obtain the best level of digestion (resulting in fragments around 36Kb in size). Ratios of *Mbo1*: dam methylase used were Tube 1.0 at 1:15,920, Tube 1.1 at 1:7,920, Tube 1.2 at 1:3,120, Tube 1.3 at 1:2,320, Tube B+ at 1:720, Tube B- at 1:720 and Tube dam containing dam methylase only (see Table 2.4 in methods). All tubes (except Tube B-) were treated with calf intestinal alkaline phosphatase to dephosphorylate the ends of the DNA fragments so preventing the formation of concatamers. The results of the digestion, along with the undigested control tubes are shown in Figure 5.1.

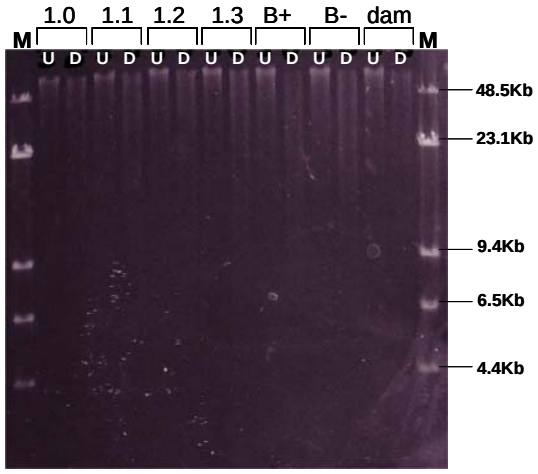


Figure 5.1 Results of competitive digestion using Mbo1 and dam methylase on whole genomic DNA from a translocation patient. U; undigested control, D; digested DNA for each enzyme dilution (1.0, 1.1, 1.2, 1.3, BASIC+, BASIC- and dam) run against M; λ DNA digested with HindIII.

The tube containing the enzyme mix showing the best digestion (in this case, tube 1.0) was selected for ligation to vector arms for production of the library. After ligation, packaging and plating, the library was incubated overnight to allow for colony growth, however, no colonies were observed.

5.2.2 EpiFOS fosmid library

5.2.2.1 Using whole genomic DNA

Whole genomic DNA was sheared by passing through a 21 gauge syringe needle five times and ten times and the fragment size assessed by gel electrophoresis (Figure 5.2).

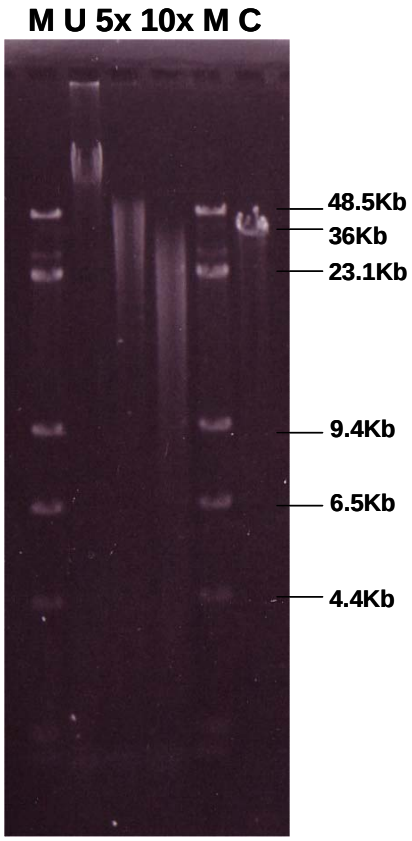


Figure 5.2 Results of shearing of whole genomic DNA from a translocation patient. U; Unsheared DNA, 5x; DNA sheared by passing 5 times through syringe needle, 10x; DNA sheared by passing 10 times through syringe needle run against M; λ DNA digested with HindIII and C; Control DNA of 36Kb.

Fragment sizes of 30 to 44Kb are required for packaging into the phage head. The gel showed that the unsheared genomic DNA was too large, and that the DNA passed 10 times through the needle was too small however the DNA passed 5 times through the needle had an average fragment size of 36Kb, the same size as the fosmid control DNA (DNA fragments at the specified size of 36Kb) which was run on the gel as a control marker. This 5 times sheared DNA was used to create the library and the packaged phage particles were plated onto agar. Clones were selected at random for FISH to assess the successful incorporation of genomic DNA into the vector. Figure 5.3 shows the FISH signal

for one such clone which was seen to hybridise to chromosome 5p15.31-5p15.33.

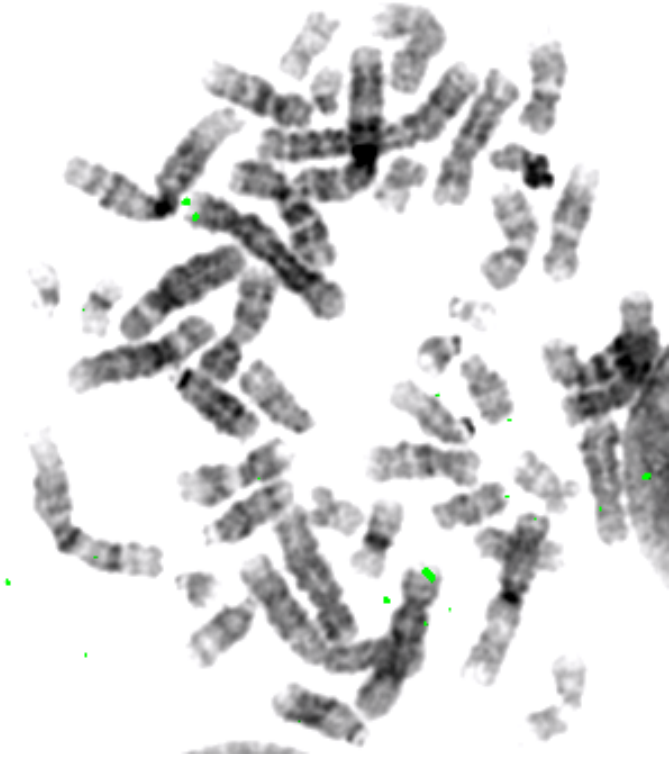


Figure 5.3 Inverted DAPI banding of metaphase chromosomes and FISH signal for a fosmid clone selected at random from the test library created using whole genomic DNA from patient t(7;13)(q31.3;q21.3). The clone hybridised to 5p15.31-5p15.33.

5.2.2.2 Using flow sorted derivative chromosomes

In order to assess the amount of starting material required to make a library suitable for screening, 2,500,000 nuclei were flow sorted and divided into 1000ng, 500ng and 250ng aliquots. Each aliquot was handled in the same way and used to generate individual libraries. Shearing of each aliquot of DNA 3 times through a syringe needle resulted in DNA fragments of approximately the same size regardless of the amount of DNA contained within the sample (Figure 5.4).

Flow sorted derivative chromosomes exhibit a lower molecular weight than isolated genomic DNA, hence reducing the amount of shearing required.

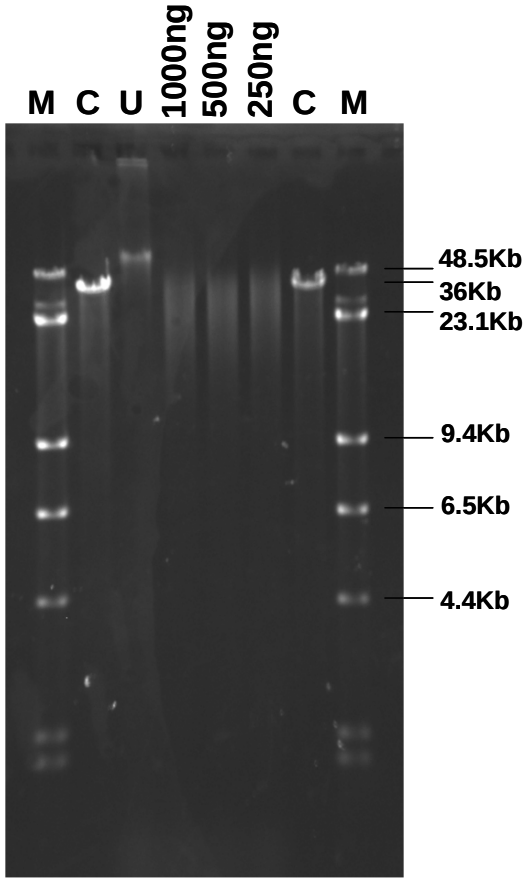


Figure 5.4 Results of shearing using starting amounts of 1000ng, 500ng and 250ng DNA. C; Control DNA of 36Kb, U; Unsheared DNA, 1000ng, 500ng and 250ng aliquots of DNA sheared by passing 3 times through syringe needle run against M; λ DNA digested with HindIII.

Each library was plated out and the number of colonies counted. The libraries generated from 1000ng starting material gave approximately 900 colonies, from 500ng gave approximately 600 colonies and from 250ng gave approximately 200 colonies. These colony numbers were obtained using 1/100th of the library suspension. Extrapolating these figures meant that the 1000ng library would generate 90,000 clones, the 500ng library would generate 60,000 clones and the

250ng library 20,000 clones. This equated to 28x coverage of the genome for the 1000ng library, 19x coverage for the 500ng library and 6x coverage for the 250ng library. Generation of a library from 250ng of starting material was selected as a method that provided reasonable coverage (approximately 3x to 18x coverage depending on the size of the derivative chromosome), whilst minimising the time required for flow sorting of the derivative chromosomes.

5.2.3 Conclusions

Use of the cosmid library protocol required 1050ng of DNA as starting material compared with the 450ng required for the fosmid library approach. This has a major impact on the time needed to flow sort the required number of derivative chromosomes (approximately 2 days for 450ng versus 5 days for 1050ng DNA).

Both library production systems utilised different methods of DNA fragmentation to reduce the molecular weight of the DNA to approximately 36Kb for packaging into the phage head; for the cosmid library, the DNA was fragmented by enzyme digestion and for the fosmid library the DNA was fragmented by shearing through a syringe needle. Size reduction of the DNA by enzyme digestion may produce a bias due to the precise nature of the fragmentation sites, whereas size reduction via shearing should be a random process and not introduce any bias.

For these reasons, and the lack of cosmid clones obtained, the fosmid library procedure was selected over the cosmid library procedure for generation of a custom-made library.

5.3 Using the EpiFOS kit to generate a library from a patient with a characterised translocation as a test case for developing screening methods

A fosmid library was created from a patient carrying a balanced reciprocal t(7;13)(q31.3;q21.3) translocation that had previously been successfully sequenced across the translocation junctions. This library was used to investigate strategies of screening for chimeric clones containing the translocation junctions. Derivative chromosomes 7 and 13 were flow sorted (Figure 5.6) and used to create two separate libraries. 450ng of each derivative chromosome (1,250,000 derivative chromosome 7 and 2,000,000 derivative chromosome 13) was flow sorted, allowing 250ng for generation of the library, and 200ng for gel electrophoresis to assess the DNA fragment size before and after shearing.

Both libraries were generated and grown on round plates as described in the methods. These plated libraries were used to generate two different set of filters for screening. Firstly, a robot was used to pick off colonies into 384 well plates and the clones gridded onto filters and secondly replica round filters were generated from the original round library plates.

The gridded filters were used to assess the methods of screening detailed in section 5.4. The round replica filters representing the whole library were used to assess the library coverage and success in generation of chimeric breakpoint spanning clones.

5.4 Comparison of screening methods using gridded library fosmid clones

A total of 14,208 clones were picked at random from the derivative chromosome 7 library and 13,824 from the derivative chromosome 13 library. These clones represented an estimated 2.8x coverage of the derivative chromosome 7 and

4.5x coverage of the derivative chromosome 13. Clones were picked into 384 well plates and gridded onto filters with a maximum of 16 plates per filter. The derivative chromosome 7 and the derivative chromosome 13 libraries were both gridded onto 3 filters each enabling efficient screening.

5.4.1 Inter Alu PCR screening

Using the translocation spanning BAC clones identified by FISH as template (RP11-384A20 for chromosome 7 and RP11-360I23 for chromosome 13) probes were generated by PCR to amplify the unique sequence present between Alu repeats within the sequence (Figure 5.5).

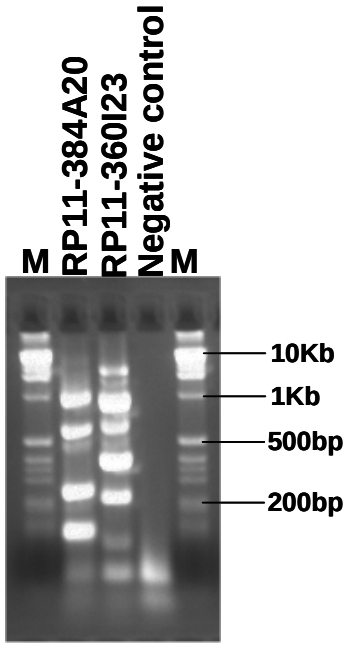


Figure 5.5 Results of Alu PCR amplification assessed by gel electrophoresis with 1Kb marker.

Hybridisation of these probes to the gridded library filters resulted in low level background, but failed to identify specific clones.

5.4.2 PCR product pools

PCR product pools were created containing specific PCR products. PCR primers were designed to amplify one product every 10Kb along the sequence of the spanning BAC clone previously identified by FISH. The chromosome 7 pool consisted of 16 products, and the chromosome 13 pool of 18 products (Appendix A4).

Hybridisation of the chromosome 7 pool of products to the derivative chromosome 7 library identified 12 clones; der7_1C5, der7_1F16, der7_15N14, der7_21H5, der7_23K18, der7_26N14, der7_27P5, der7_35D10, der7_35F10, der7_37B6, der7_37H21 and der7_37N12. Screening the same library with the chromosome 13 pool of products identified 5 clones; der7_2I21, der7_8N13, der7_12B20, der7_28J18 and der7_37N12. One clone; der7_37N12 was identified by both pools implying that the clone contained both chromosome 7 and chromosome 13 DNA and potentially spanned the derivative chromosome breakpoint.

Hybridisation of the chromosome 13 pool of products to the derivative chromosome 13 library identified 2 clones; der13_34D21 and der13_36M16, however, the chromosome 7 pool did not identify any clones. The low numbers of clones identified during the screening suggested that the coverage of this library is poor around the breakpoint region. It was unlikely that a derivative chromosome 13 junction spanning clone is contained within this gridded library as the pools failed to identify a common clone.

5.4.3 End sequencing

A total of 28,032 colonies were picked by the robot from the round agar plates into 384 well plates for gridding and end sequencing; 14,208 from the derivative chromosome 7 library and 13,824 from the derivative chromosome 13 library. Of these, 384 failed to grow; 244 from the derivative chromosome 7 library and 140

from the derivative chromosome 13 library. The remaining clones (13,964 derivative chromosome 7 clones and 13,864 derivative chromosome 13 clones) were end sequenced with a high success rate (Table 5.1) and the resulting data compared with the human reference sequence to investigate the origin of the insert clone DNA with the ultimate aim of finding clones chimeric for chromosome 7 and 13 sequence, and hence containing the translocation breakpoints.

Number of clones	der7	der13
Submitted for end sequencing	13,964	13,684
Passed both ends	12,460 (89.2%)	10,009 (73.1%)
Passed one end only	989 (7.1%)	1,426 (10.4%)
Failed both ends	515 (3.7%)	2,249 (16.5%)

Table 5.1 Summary of end sequencing reads obtained from derivative chromosome 7 and derivative chromosome 13 library clones.

All clones which produced sequence greater than 20bp in length were compared with the human reference sequence to align the clone sequence to it. 88.9% (11,081/12,460) of clones from the derivative chromosome 7 library and 84.2% (8,428/10,009) of clones from the derivative chromosome 13 library were successfully aligned to the human reference sequence (Table 5.2).

der7	der13
6550 (59.1%) on chr7	3600 (42.7%) on chr7
2225 (20.1%) on chr13	4201 (49.8%) on chr13
2022 (18.2%) on chr6	249 (3.0%) on chr14
265 (2.4%) on others	368 (4.4%) on others

Table 5.2 Summary of fosmid insert DNA alignments against the human genome reference sequence for the derivative chromosome 7 and derivative chromosome 13 library clones.

In the derivative chromosome 7 library, 79.2% of clones hit chromosomes 7 or 13 as expected. A further 18.2% aligned to chromosome 6. For the derivative chromosome 13 library, 92.5% of clones aligned to chromosomes 7 or 13 as expected and a further 3.0% aligned to chromosome 14. Analysis of the patient's

flow karyogram revealed that the derivative chromosome 7 lay close to chromosome 6 and the derivative chromosome 13 lay close to chromosome 14 (Figure 5.6) explaining the contamination seen in the libraries. During the flow sorting process, a small number of chromosomes can become fragmented. This DNA, once carried through the library preparation procedure would explain the fosmid clones that show DNA inserts that aligned to other chromosomes.

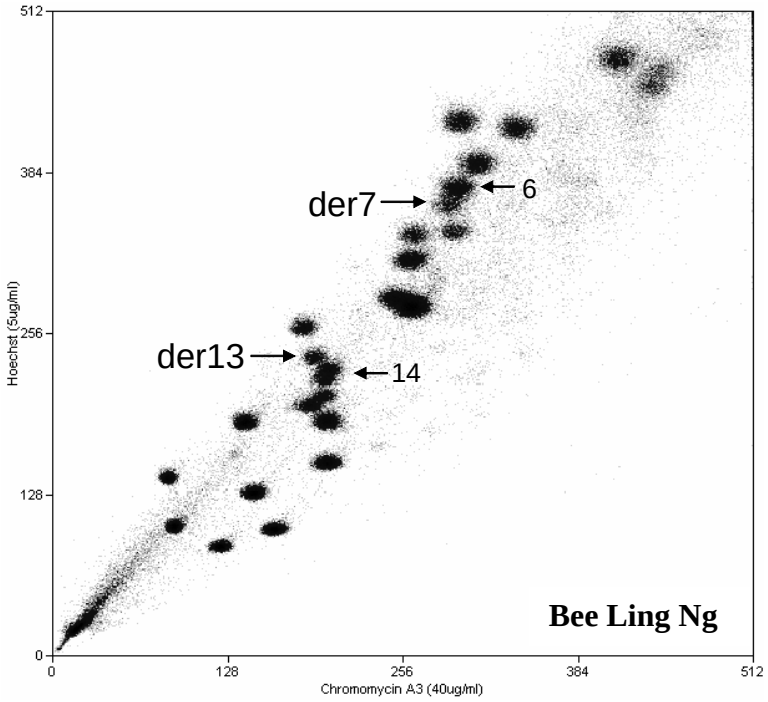


Figure 5.6 Flow karyogram for patient t(7;13)(q31.3;q21.3). The derivative chromosome 7 was close to the normal chromosome 6 and the derivative chromosome 13 close to the normal chromosome 14.

19 clones from the derivative chromosome 7 library and 10 clones from the derivative chromosome 13 library could not be aligned to the human reference sequence because the two ends of each clone aligned to different chromosomes. Further analysis of the end sequences revealed that all 29 of these clones were chimeric for both chromosome 7 and chromosome 13 sequence. One clone

(der7_37N12) was chimeric for chromosome 7 and chromosome 13 sequences at the correct locations, and therefore potentially contained the translocation junction sequence for the derivative chromosome 7. The remaining 28 clones which mapped to unexpected regions of chromosomes 7 and 13 must have been produced as a by-product of the fosmid library construction; incomplete phosphorylation of the ends of the insert DNA could have allowed smaller fragments of DNA to join, resulting in an artificially created fragment of DNA that is the correct size to be packaged into the phage head.

5.4.4 Specific breakpoint PCR product

In order to assess the number of chimeric fosmid clones present in the gridded libraries, two PCR products which spanned the derivative chromosome 7 and 13 translocation junctions were created (Table 5.3).

Primer	Forward primer	NCBI Build 35 position	Reverse primer	NCBI Build 35 position
der7	GTAGTGATTCGGCCTTGCAT	chr7:121245417-121245436	TGGCCATATTTGGCTTTTTG	chr13:71023936-71023955
der13	TCCATTCATGTTGCTGCATT	chr13:71023730-71023749	GGAAGACAGGATGGATTCAAA	chr7:121245470-121245490

Table 5.3 PCR primers used to create derivative chromosome specific products for library screening.

Screening of the derivative 7 library with the derivative 7 specific breakpoint PCR product identified 3 fosmid clones; der7_12B20, der7_28J18 and der7_37N12. Screening of the derivative 13 library with the derivative 13 specific breakpoint PCR product identified a single fosmid clone; der13_10O8. Analysis of the end sequence data previously obtained for these clones showed that der7_37N12 was chimeric for chromosome 7 and 13 sequence and contained the derivative chromosome 7 translocation junction. However, the other three fosmid clones aligned to regions of the genome not believed to be involved in the translocation (Table 5.5). The insert sizes for these clones were 41,206bp for der7_12B20, 36,062bp for der7_28J18 and 34,447bp for der13_10O8. These clones were subsequently analysed by FISH.

Fosmid clone der7_12B20 hybridised mainly to chromosome 7q21.2-7q31.3 with weaker signals on 13q21.31-13q31.2 and 7p12.1-7p14.2. Analysis of the end sequence data showed that the clone localised to 7p12.1 of the human genome reference sequence. These results suggested that the well contained mixed colonies due to incomplete separation of phage clones during the plating out stage of library production.

FISH analysis of der7_28J18 showed a hybridisation signal at 6q21-6q22.33 whilst the end sequence data aligned the clone to 6q22.1. Blast analysis of the sequence for this region showed no homology to the specific breakpoint product used for library screening, suggesting that the well must contain a mixture of 2 clones from poor separation during plating.

Analysis of der13_10O8 showed alignment to 13q13.1 by end sequence analysis and hybridisation to 13q12.3-13q14.13 and 13q14.3-13q21.3 by FISH. This FISH data showed that the clone lay close to or at the translocation breakpoint. This region may have been sufficient for the breakpoint specific PCR product used during radioactive screening to hybridise to, producing a positive result. The more proximal FISH signal may have been the result of a mixed well.

In order to confirm that these 3 clones were from mixed colonies, each well would need to be streaked to single colony and re-analysed.

5.4.5 Summary of fosmid clones identified by the different screening methods

A summary of the clones identified by the different screening strategies is depicted in Table 5.4.

		derivative chromsome 7 library	derivative chromsome 13 library
PCR pool screening	chr7 pool	der7_1C5, der7_1F16, der7_15N14, der7_21H5, der7_23K18, der7_26N14, der7_27P5, der7_35D10, der7_37B6, der7_37H21, der7_37N12	None
	chr13 pool	der7_2I21, der7_8N13, der7_12B20, der7_28J18, der7_37N12	der13_34D21, der13_36M16
Specific breakpoint product		der7_12B20, der7_28J18, der7_37N12	der13_10O8

Table 5.4 Summary of fosmid clones identified by PCR pool screening and screening with specific breakpoint products. der7_37N12 was identified by both approaches as containing the derivative chromosome 7 translocation junction.

For the derivative chromosome 7 library, screening by PCR product pools and using a specific breakpoint product identified clone der7_37N12 which had previously been identified by end sequencing to contain the derivative chromosome 7 translocation junction sequence.

None of the screening strategies identified a chimeric clone in the derivative chromosome 13 library. A likely explanation for this is that the gridded library did not actually contain a derivative chromosome 13 spanning fosmid.

The end sequence data for all clones detailed in Table 5.4 was retrieved and summarised in Table 5.5. All clones identified by PCR product pools were seen to align close to the translocation breakpoints. These clones were identified because the PCR products used to screen the library were designed at a 10Kb resolution across the whole length of the spanning BAC clone.

Clonename	Primer	BLAT hit (May 2004)		
		chr	start	end
der7_1C5	Forward	no data		
	Reverse	no data		
der7_1F16	Forward	7	121,159,146	121,159,888
	Reverse	7	121,195,221	121,194,772
der7_2I21	Forward	13	70,922,368	70,923,133
	Reverse	13	70,958,461	70,957,741
der7_8N13	Forward	13	71,085,684	71,086,475
	Reverse	13	71,123,862	71,123,035
der7_12B20	Forward	7	51,986,318	51,986,666
	Reverse	7	51,944,751	51,945,460
der7_15N14	Forward	no data		
	Reverse	no data		
der7_21H5	Forward	7	121,199,488	121,200,334
	Reverse	7	121,236,279	121,235,430
der7_23K18	Forward	7	121,195,574	121,196,211
	Reverse	7	121,229,567	121,228,758
der7_26N14	Forward	no data		
	Reverse	7	121,154,377	121,155,087
der7_27P5	Forward	7	121,118,879	121,185,690
	Reverse	7	121,221,332	121,220,529
der7_28J18	Forward	6	117,769,406	117,770,080
	Reverse	6	117,805,731	117,806,142
der7_35D10	Forward	7	121,183,478	121,184,256
	Reverse	7	121,223,748	121,223,501
der7_35F10	Forward	7	121,129,871	121,130,669
	Reverse	7	121,165,835	121,165,011
der7_37B6	Forward	7	121,160,129	121,160,933
	Reverse	7	121,198,097	121,197,251
der7_37H21	Forward	7	121,193,308	121,193,601
	Reverse	no data		
der7_37N12	Forward	13	71,035,495	71,034,670
	Reverse	7	121,226,143	121,226,939
der13_10O8	Forward	13	32,131,659	32,132,406
	Reverse	13	32,166,106	32,165,301
der13_34D21	Forward	13	70,922,561	70,923,411
	Reverse	13	70,955,023	70,954,228
der13_36M16	Forward	13	70,959,057	70,959,865
	Reverse	13	70,994,435	70,993,649

Table 5.5 End sequence data available for all clones identified by alternative screening strategies.

5.4.6 Conclusions

In total, 14,208 colonies were picked from the derivative chromosome 7 agar plates, resulting in an estimated 2.8x coverage along the chromosome. For derivative chromosome 13, 13,824 colonies were picked resulting in an estimated 4.5x coverage. These gridded clones were used to generate filters used to establish a method of screening custom-made translocation libraries.

End sequencing of every clone within the gridded libraries identified a potential spanning clone for the derivative chromosome 7 library, but not the derivative chromosome 13 library. However, it was possible to use the data obtained from the end sequencing of all clones to refine the mapping of the translocation breakpoint region on chromosome 13 where a spanning clone was not identified. Analysis of the alignment of the clones from both libraries to the human reference sequence along chromosome 13 revealed a switch from one chromosome to the other (Figure 5.7) so indicating the breakpoint region. The different depths of coverage for each library (2.8x for the derivative chromosome 7 library and 4.5x for the derivative chromosome 13 library) are also illustrated in Figure 5.7.

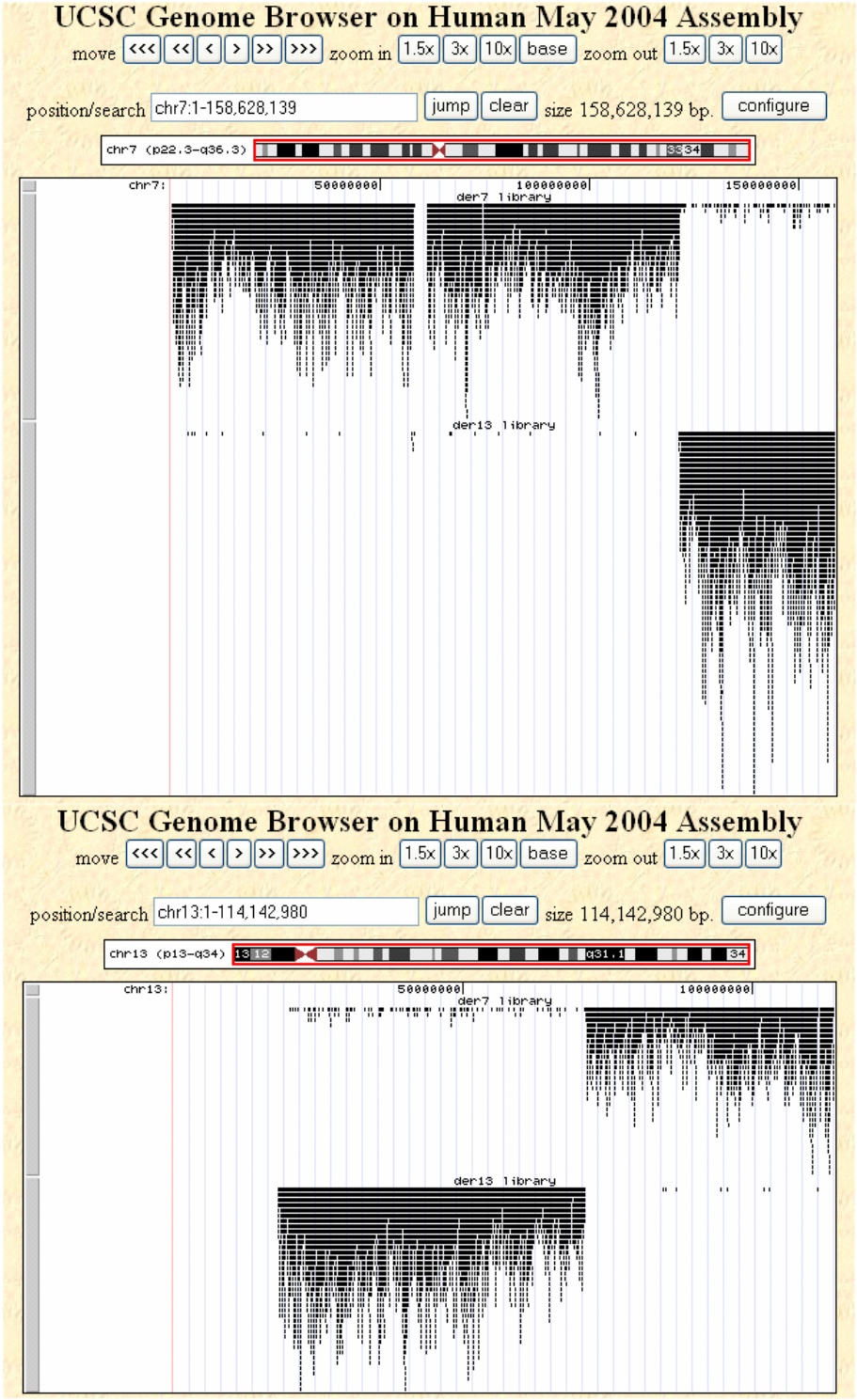


Figure 5.7 UCSC download showing clone coverage across a; chromosome 7 for both derivative chromosome 7 and 13 libraries and b; chromosome 13 for both derivative chromosome 7 and 13 libraries.

In cases where no spanning clone was identified, as in the case of the derivative chromosome 13, the translocation junction could be mapped between the clones mapping to chromosome 13 in the derivative chromosome 13 library and the clones mapping to chromosome 13 in the derivative chromosome 7 library (Figure 5.8). The breakpoint was mapped between clones 3615_13_36m16 from the derivative chromosome 13 library and Der7_8n13 from the derivative chromosome 7 library. The chromosome 13 breakpoint region that fell between these clones was from 70,966,023 to 71,085,684bp, a gap of 130,611bp. Further mapping by STS PCR would have been required in this situation to refine the breakpoint region prior to LR PCR.

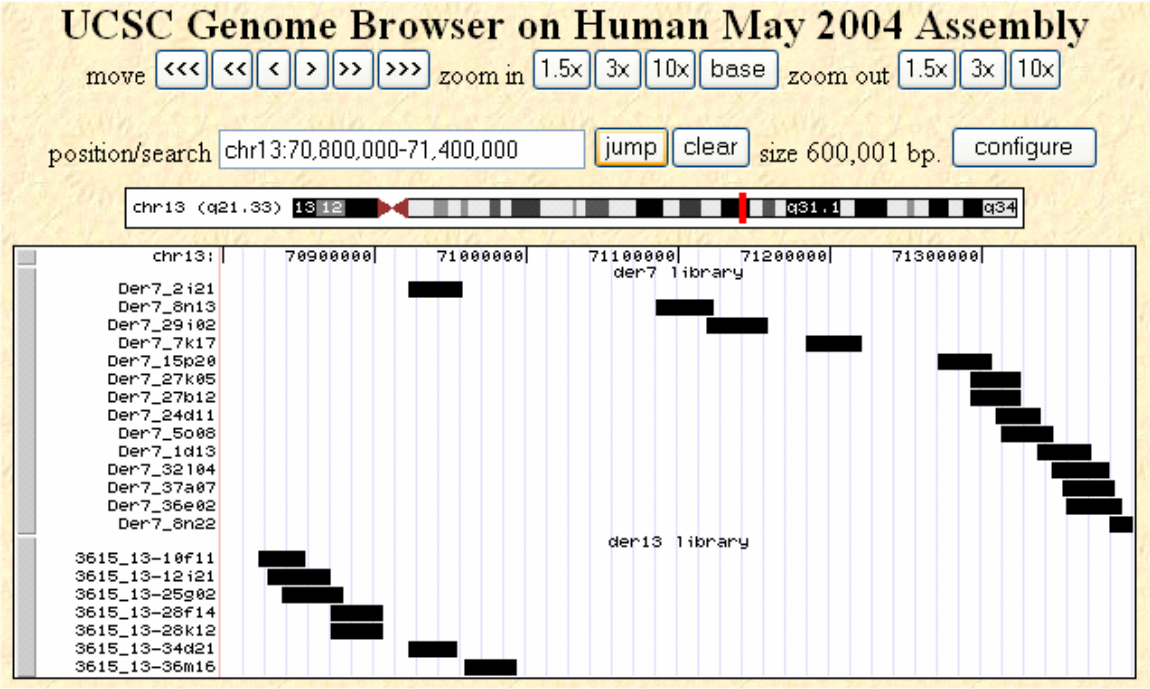


Figure 5.8 Fosmid clone coverage across the breakpoint region of chromosome 13 for the derivative chromosome 7 and 13 libraries. No spanning clone was identified by end sequencing across this breakpoint. Clone der7_2i21 from the derivative chromosome 7 library aligned proximal to the chromosome 13 breakpoint. This clone must have been generated from a chromosome 13 flow sorting artefact as the derivative chromosome 13 and normal chromosome 13 peaks were close to each other on the flow karyogram (Figure 5.6).

Screening the libraries by end sequencing and PCR product pools both identified the derivative chromosome 7 breakpoint spanning clone. The data generated by end sequencing provided a valuable resource for assessing and verifying results obtained by screening with PCR product pools. In this instance the end sequence data confirmed the simplicity of the rearrangement but in patients with complex rearrangements the end sequence data may help to resolve the rearrangement structure. However, end sequencing of an entire library remains an expensive process, so may not be a viable option for routine breakpoint mapping. Screening by PCR product pools involves the hazards associated with radiation but remains a faster and less costly method than screening by end sequencing.

5.4.7 Screening of round library plates

Screening of the derivative chromosome 7 gridded library with 2.8x coverage by end sequencing found 1 translocation junction spanning clone but the derivative chromosome 13 library with 4.5x coverage did not reveal a translocation spanning clone.

The clones in the gridded version of the library were picked by robots which could only select clones that were discreet, circular and larger than 1.5mm in diameter. For this reason, the gridded library did not contain all the clones originally created and plated out. To investigate whether translocation spanning clones were present on the round library plates, but had failed to be selected by the robots, lift filters were created from the original plated libraries for screening.

The same breakpoint specific products used to screen the gridded filters were used to screen the round filters. Due to the construction of the robot, the plates were not picked in a sterile environment, and unfortunately some plates were lost due to infection. The derivative chromosome 7 library was originally plated onto 115 plates, and subsequently 39 were unusable by the time it came to make the

filters. For the derivative chromosome 13 library, 31 of 87 plates were not available to be used for lift filter generation.

Screening of the derivative chromosome 7 library plates identified 2 clones, der7_round76 and der7_round78. Screening of the derivative chromosome 13 library plates identified 3 clones, der13_round14, der13_round60 and der13_round80. To verify that the identified clones contained the translocation junction, the clones were streaked to single colony, 10 colonies were selected and colony PCR using primers designed to amplify across the translocation junctions was used to verify that the translocation junction was present (Table 5.3). For the derivative chromosome 7 clones, der7_round76 could not be confirmed as containing the translocation junction, but der7_round78 was positive by colony PCR. For the 3 derivative chromosome 13 clones identified; der13_round 14 failed to grow, der13_round60 was confirmed by colony PCR and the plate containing der13_round80 exhibited infection, so could not be accessed for confirmation. Fosmid der13_round60 was subsequently sequenced in full and analysis of the sequence revealed that the clone sequence was identical across the derivative chromosome 13 translocation junction as had previously been identified by array painting and PCR (see Figure 3.29).

The screening of the round library filters showed that while the DNA fragments containing the derivative chromosome junction sequences were cloned during the library process, they were not necessarily selected by the robots for gridding and that as a result, the coverage of the gridded library was lower than the coverage actually generated.

Generation of the gridded library filters was a less time consuming process than generation of the round lift filters and screening of those gridded libraries also greatly reduced the amount of radioactive isotope required during the screening process. However, not all clones created in the library were selected for gridding,

emphasising the need to retain the original round library plates as a back-up if screening of the gridded library failed to identify chimeric clones.

5.5 Mapping translocation breakpoints using a custom-made fosmid library from a patient with a t(2;6)(q21.1;q25.1) translocation

To test the method of translocation breakpoint mapping using custom-made fosmid libraries, a patient with an uncharacterised reciprocal translocation was selected. This patient had not undergone any mapping other than routine G-banding which identified a t(2;6)(q21.1;q25.1) translocation (Figure 5.9). To map the translocation breakpoints prior to generation of a PCR product pool for screening of the library, flow sorted derivative chromosomes were array painted onto a WGTP microarray.

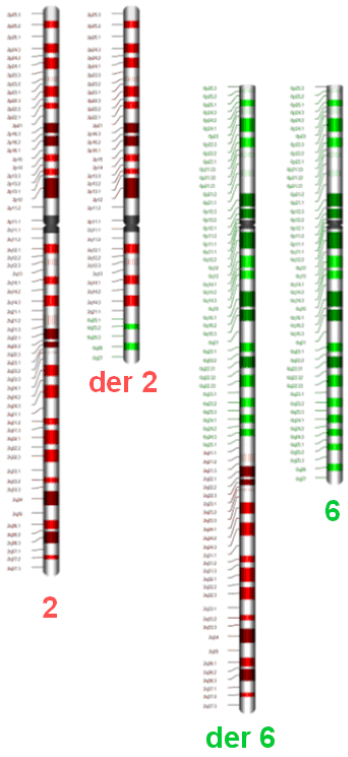


Figure 5.9 Ideogram images depicting normal chromosomes 2 and 6 and derivative chromosomes 2 (der 2) and 6 (der 6) in a patient with a t(2;6)(q21.1;q25.1) translocation.

5.5.1 Isolation of derivative chromosomes by flow sorting

Flow sorted derivative chromosomes were required to map the translocation breakpoints by array painting and for generation of the custom-made fosmid library. Assessment of the patient's karyotype (Figure 5.10) revealed that both derivative chromosomes were different from the normal homologues in size and base composition and therefore could be segregated by flow sorting. The expected size of the derivative chromosome 2 was approximately 160Mb and the derivative chromosome 6 approximately 287Mb.

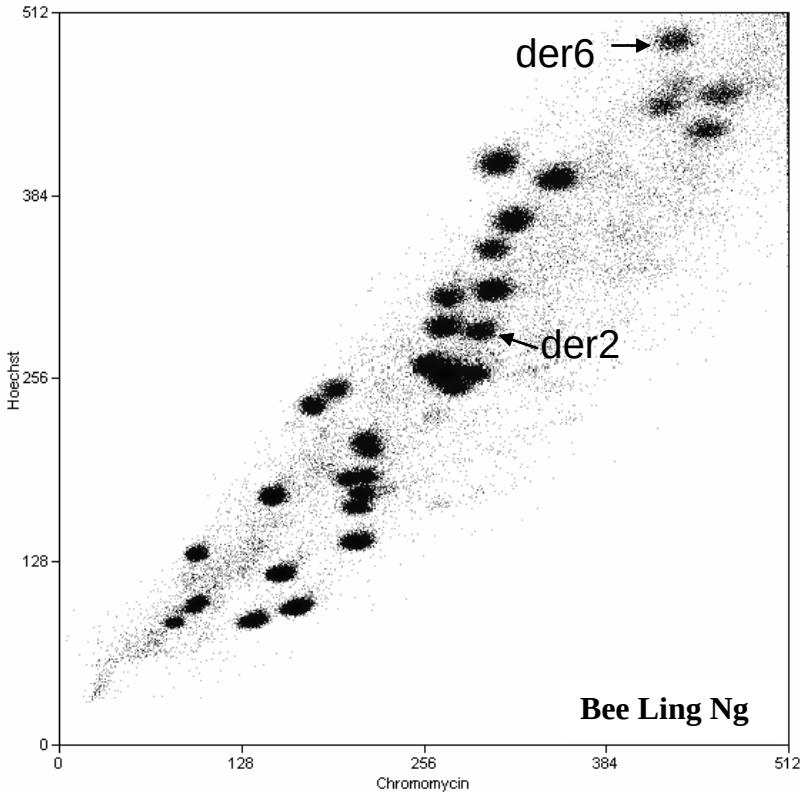


Figure 5.10 Flow karyogram for patient t(2;6)(q21.1;q25.1).

5.5.2 Verification of flow sorted derivative chromosomes by reverse chromosome painting.

The flow sorted derivative chromosomes were reverse painted onto metaphase chromosomes from a normal cell line as a method of verification and to assess the amount of contamination from other chromosomes (Figure 5.11).

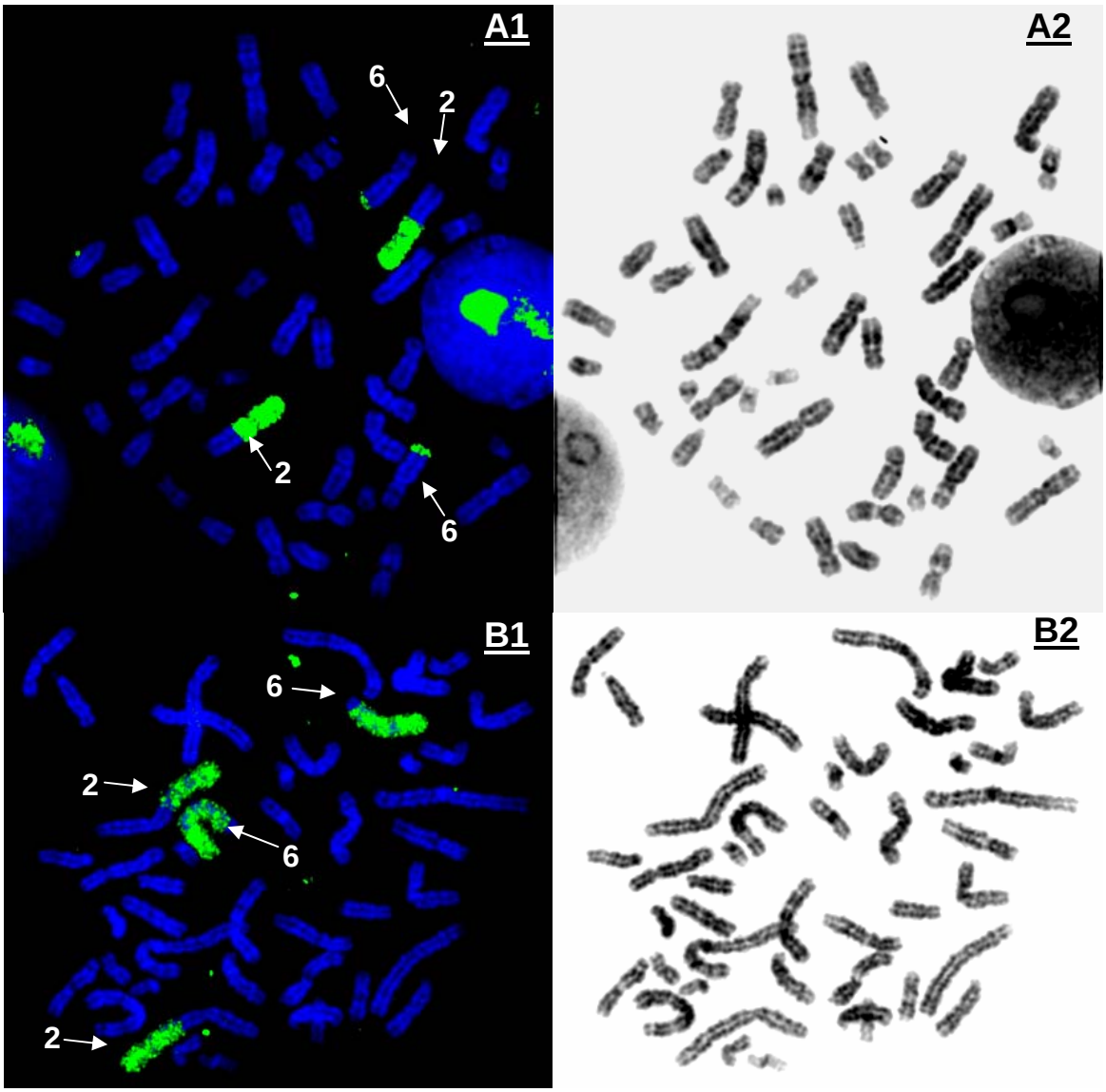


Figure 5.11 Reverse chromosome painting results for A flow sorted derivative chromosome 2 and B derivative chromosome 6. 1 metaphase FISH image and 2 inverted DAPI banding of metaphase chromosomes.

Reverse chromosome painting showed that the flow sorted derivative chromosome 2 hybridised to 2p and part of 2q and also the telomeric portion of chromosome 6q. The flow sorted derivative chromosome 6 hybridised to 6p and the majority of 6q and also the distal two thirds of 2q. Neither hybridisation detected any signal on chromosomes other than 2 and 6, suggesting that at the resolution afforded by metaphase chromosomes, the rearrangement was a simple reciprocal translocation, that no other chromosomes were involved and that the sort was not significantly contaminated.

5.5.3 Mapping translocation breakpoints by array painting

To refine the breakpoint regions from the cytogenetic bands assigned by G-banding analysis, the derivative chromosomes were differentially labelled and hybridised in an array painting experiment to a the Sanger 30K Whole Genome Tiling Path microarray (Figure 5.12).

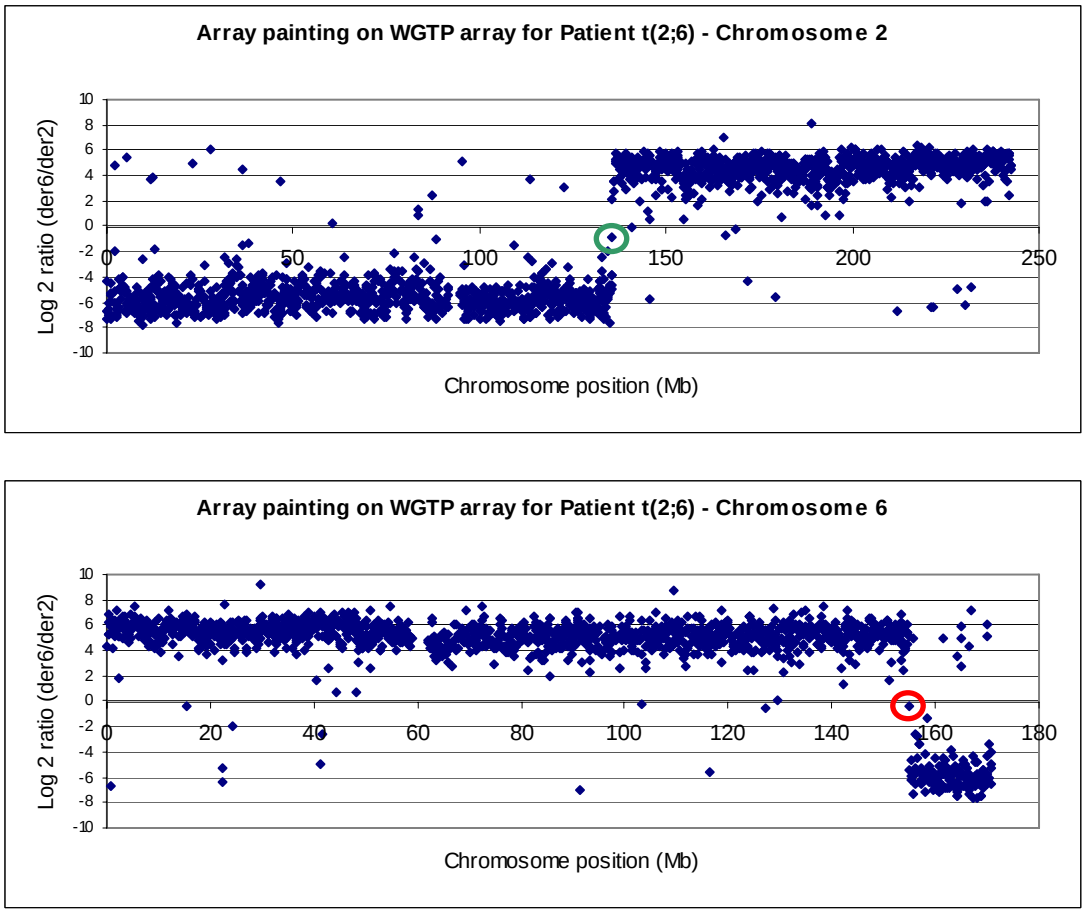


Figure 5.12 Array painting results for patient t(2;6)(q21.1;q25.1) on the whole genome tilepath microarray. Spanning clones; RP11-341H1 and RP11-486M3.

Analysis of the clone ratios in the array painting experiment revealed a single spanning clone on chromosome 2; RP11-341H1 and a single spanning clone on chromosome 6; RP11-486M3. Both spanning clones exhibited ratios close to zero implying that the translocation breakpoints lay towards the centre of the genomic DNA insert contained within the clone. Identification of the breakpoint regions at this resolution enabled the creation of PCR products pools for the screening of a custom-made library.

5.5.4 Isolation of breakpoint spanning fosmid clones using a custom-made library

5.5.4.1 Production of the t(2;6)(q21.1;q25.1) fosmid library

Approximately 450ng of both derivative chromosomes was flow sorted (corresponding to 1,750,000 derivative 2 chromosomes and 1,000,000 derivative 6 chromosomes) for generation of the derivative chromosome fosmid libraries. 450ng of chromosomes from the 9-12 peak was also sorted to be used as a control. The DNA was sheared by passing 3 times through a syringe needle (Figure 5.13).

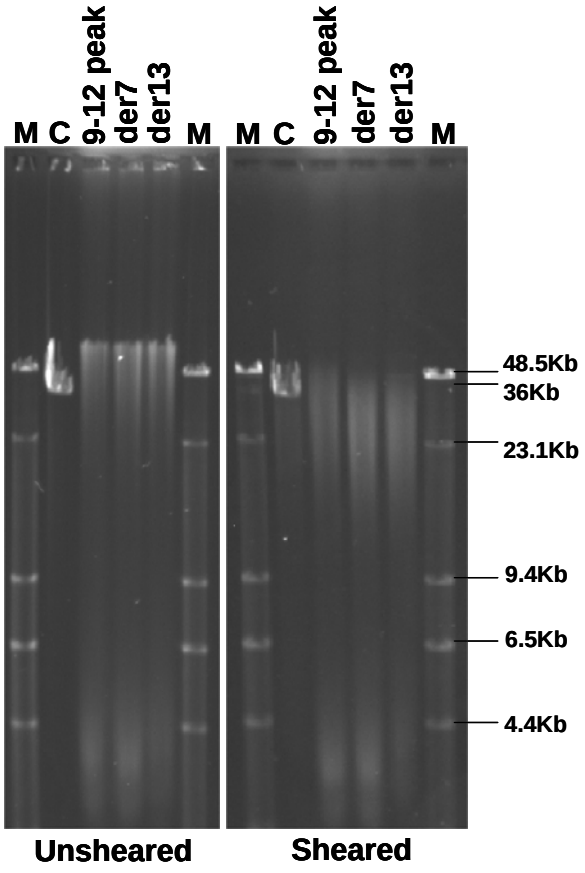


Figure 5.13 Results of DNA shearing of flow sorted chromosomes 9-12, derivative chromosome 2 and derivative chromosome 6 DNA for patient t(2;6)(q21.1;q25.1) run against M; λ DNA digested with HindIII and C; control DNA of 36Kb.

5.5.4.2 Screening of the t(2;6)(q21.1;q25.1) library using PCR product pools.

The sequence associated with both spanning BAC clones identified by the WGTP array paint experiment (Section 5.5.3) was used to design the pools of PCR products for screening of the custom-made fosmid library. The chromosome 2 spanning BAC, RP11-341H1 (Accession AC016725.5) had 162,180bp of sequence associated with it from 135,311,916 to 135,474,095bp. The chromosome 6 spanning BAC, RP11-486M3 (Accession AL357075.17) was 178,461bp in length from 154,757,465 to 154,935,925bp. The extended regions selected for primer design were from 135,311,000 to 135,475,000bp on chromosome 2 and 154,756,000 to 154,937,000bp on chromosome 6 which allowed for the possibility that the full insert of the clone had not been sequenced.

A total of 19,200 clones (4.3x coverage) of the derivative chromosome 2 library and 19,968 clones (2.5x coverage) of the derivative chromosome 6 library were picked by robots and gridded onto filters for screening.

Primer pairs were designed to generate 1 product every 10Kb along the selected regions, resulting in a pool of 17 products covering the chromosome 2 sequence and 19 products covering the chromosome 6 sequence (Appendix A5). Use of these primer pools to screen both the derivative chromosome 2 and 6 libraries revealed 3 fosmid clones identified by both pools for the derivative chromosome 2 library; der2_16D21, der2_33M22 and der2_42P23 and 2 fosmid clones identified by both pools in the derivative chromosome 6 library; der6_9G19 and der6_35N15. All 5 fosmid clones were end sequenced to confirm that they were chimeric for the translocation (Figure 5.14).

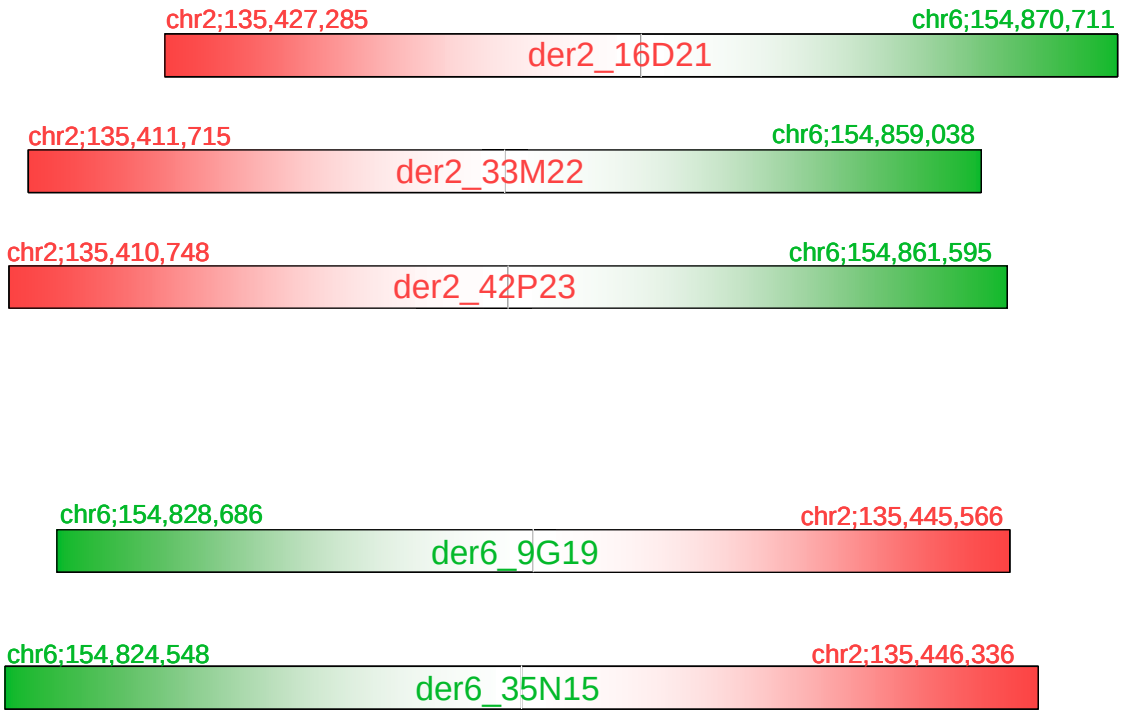


Figure 5.14 Alignment of chimeric fosmid clones by end sequence data.

On the basis of the end sequence data obtained from all 5 chimeric fosmid clones, der2_42P23 and der6_35N15 were selected for full sequencing as they appeared to contain the translocation junctions towards the middle of the clone DNA.

5.5.5 Identification of translocation breakpoints by sequence analysis

Analysis of the sequence from clone der2_42P23 showed that the insert DNA aligned to chromosome 2 from 135,410,748 to 135,432,727bp and chromosome 6 from 154,849,221 to 154,861,595bp and that clone der6_35N15 aligned from 154,824,548 to 154,849,219bp of chromosome 6 and 135,432,735 to 135,446,336bp of chromosome 2. The chromosome breakpoints were therefore deemed to be at 135,432,727bp on chromosome 2 and 154,849,219bp on chromosome 6.

Analysis of the sequence around the breakpoints revealed a 7bp deletion from chromosome 2 and a 3bp deletion from chromosome 6 (Figure 5.15).

```
t(2;6)(q21.1;q25.1)
chr2; GCTTTTGTTAGTTTCCCTGGAGCATCTATTTTTGTAATACATTTTGCCAGTGGGACTGAAAAGCT
der2; GCTTTTGTTAGTTTCCCTGGA          cagtgatttctcctgcctcggcctcctgggtagc
der6; cggctcatggcagcctccgcctcctggcTTTTTGTAATACATTTTGCCAGTGGGACTGAAAAGCT
chr6; cggctcatggcagcctccgcctcctggcttccagtgatttctcctgcctcggcctcctgggtagc
```

Figure 5.15 Alignment of sequence across the translocation junctions to the human genome reference sequence for patient t(2;6)(q21.1;q25.1). Deleted bases are shown in red.

Analysis of the full sequence generated from both der2_42P23 and der6_35N15 showed that the sequence aligned to the human genome reference sequence with 100% homology, indicating that there was no additional complexity.

5.6 Conclusions

Studies have shown that apparently balanced reciprocal translocations sometimes exhibit more complexity than originally observed through G-banding. If this additional complexity is around the translocation breakpoint, mapping by array painting and PCR may not be sufficient to fully resolve the translocation junctions. Conventional methods of breakpoint mapping such as FISH or generation of somatic cell hybrids followed by Southern blotting can be laborious, enhancing the need for a robust method of mapping these more complicated breakpoints. This chapter describes how the generation of a custom-made fosmid library and its screening using PCR product pools is a viable method for the identification of clones containing the translocation junctions. Once a spanning BAC clone has been identified by array painting or FISH, generation and screening of the custom-made library is sufficient for the isolation of the breakpoint sequence. Sequencing of the identified spanning fosmid clones generates approximately 36Kb of sequence around the translocation, providing information about the sequence surrounding the breakpoints. This analysis would

reveal any additional complexity around the translocation breakpoints which may have proven difficult to map by array painting and PCR.

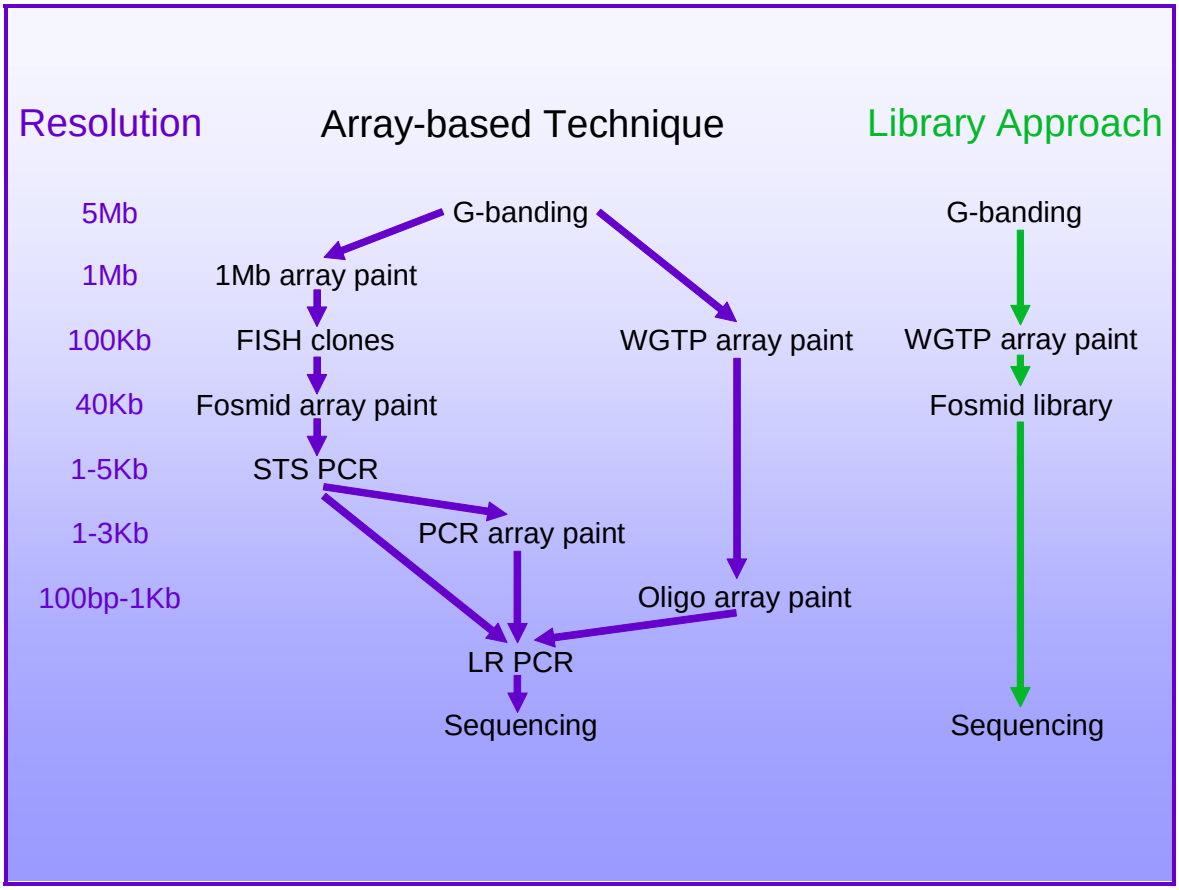


Figure 5.16 Comparison between array-based methodologies and custom-made libraries for breakpoint mapping.

A comparison between breakpoint mapping by array painting and PCR and custom-made fosmid libraries is shown in Figure 5.16. Use of custom-made fosmid libraries for breakpoint mapping does not rely on the assumption that a reciprocal translocation is balanced, and generates sequence data around the breakpoint which can reveal any cryptic rearrangements that may be present.

6 Results: Investigation of viral integration using microarrays and custom-made libraries

6.1 Introduction

The human herpes virus is ubiquitous in the human population and normally exists as an extra chromosomal element in the host. HHV-6 is a lymphotropic virus which has been identified as a causative agent in the illness exanthum subitum (also known as Roseola infantum) whose patients present with rash, fever and febrile convulsions. Restriction analysis and FISH studies have shown that in approximately 3% of the British population the HHV-6 virus has integrated into the human genome (Luppi et al. 1993; Torelli et al. 1995; Morris et al. 1999).

This chapter demonstrates how the techniques developed to map chromosome rearrangement breakpoints can be applied to genomic rearrangements such as the integration of viral genomes into the human genome. FISH onto metaphase chromosomes was used to identify the chromosomes with viral integration, microarrays were used to assess the viral copy number within the human genome, and a custom-made library was used to investigate the composition of the integrated virus. These techniques, along with FISH onto extended chromatin fibres have generated data that may help to elucidate the mechanism by which the HHV-6 virus integrates into the human genome.

6.1.1 The HHV-6 virus structure

The HHV-6 virus was extensively mapped by restriction enzyme digestion using a combination of 5 different enzymes (Figure 6.1) (Lindquester et al. 1996). The virus consists of a central unique segment of approximately 144.5Kb in length, flanked by 2 direct repeats either side (DR_L and DR_R). Whole genomic DNA isolated from the virus and cloned fragments covering approximately 103.9Kb (72%) of the unique segment were obtained for use in this study from Duncan Clark at University College London and Scott Schmid at CDC, Atlanta.

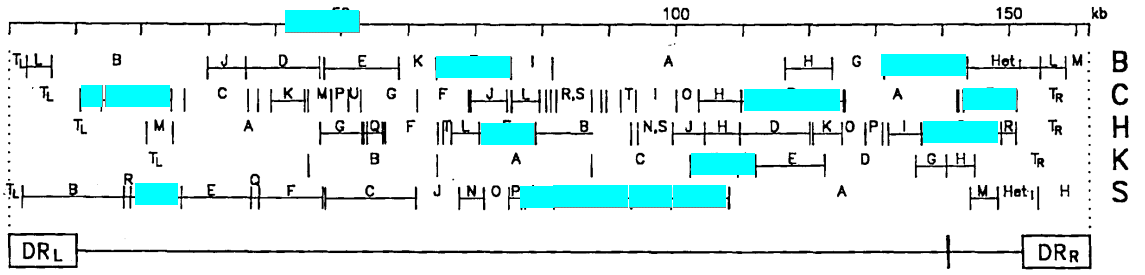


Figure 6.1 Restriction digest map of the HHV-6B genome using enzymes B; BamHI, C; ClaI, H; HindIII, K; KpnI and S; SalI (taken from Lindquester et al. 1996). Cloned fragments (blue) were obtained for use in this study.

6.1.2 Patient details

Lymphoblastoid cell-lines were obtained from 3 patients (AMD, AS and KK) suspected of containing integrated HHV-6 after HHV-6 PCR testing.

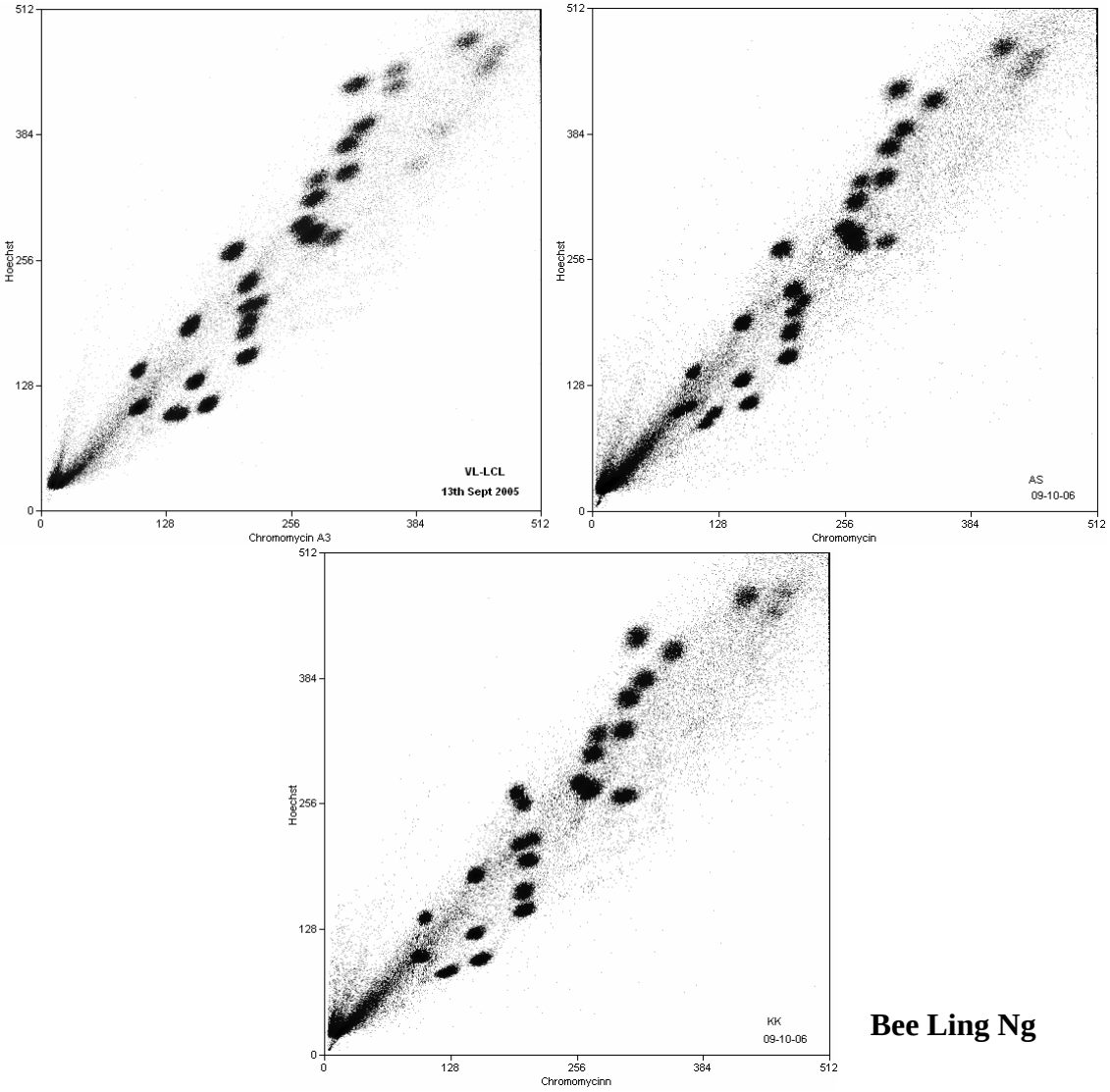
Patient AMD was a healthy 35 year old male with no known past medical history. HHV-6 specific quantitative competitive PCR had established that a high viral load equating to approximately 5 copies per cell was retained by the patient over a 10 month period (Clark et al. 1996). PCR primers used to amplify each of the 87 predicted open reading frames of the virus showed that all were present.

Patient AS was a 24 year old male who had previously developed chronic renal failure and received a renal transplant at the age of 18 years. Following the transplant, the patient presented with lethargy and severe weight loss. PCR tests for HHV-6 showed a raised level of $12x10^6$ copies/ml, with a subsequent sample 8 months later showing a persistently high level of $8x10^6$ copies/ml.

Patient KK was a 25 year old male who developed chronic renal failure and received a cadaveric renal transplant aged 19 years. Following a bout of lethargy, HHV-6 PCR testing revealed a viral load of $8.6x10^6$ copies/ml and a follow-up sample 2.5 years later showed a level of $12x10^6$ copies/ml.

6.1.2.1 Flow karyograms

The flow karyograms were assessed for all 3 patients (Figure 6.2), however in all cases there was no notable change in the profile when compared to a karyogram from a normal individual, therefore no indication was given of which chromosome contained the integrated viral DNA.

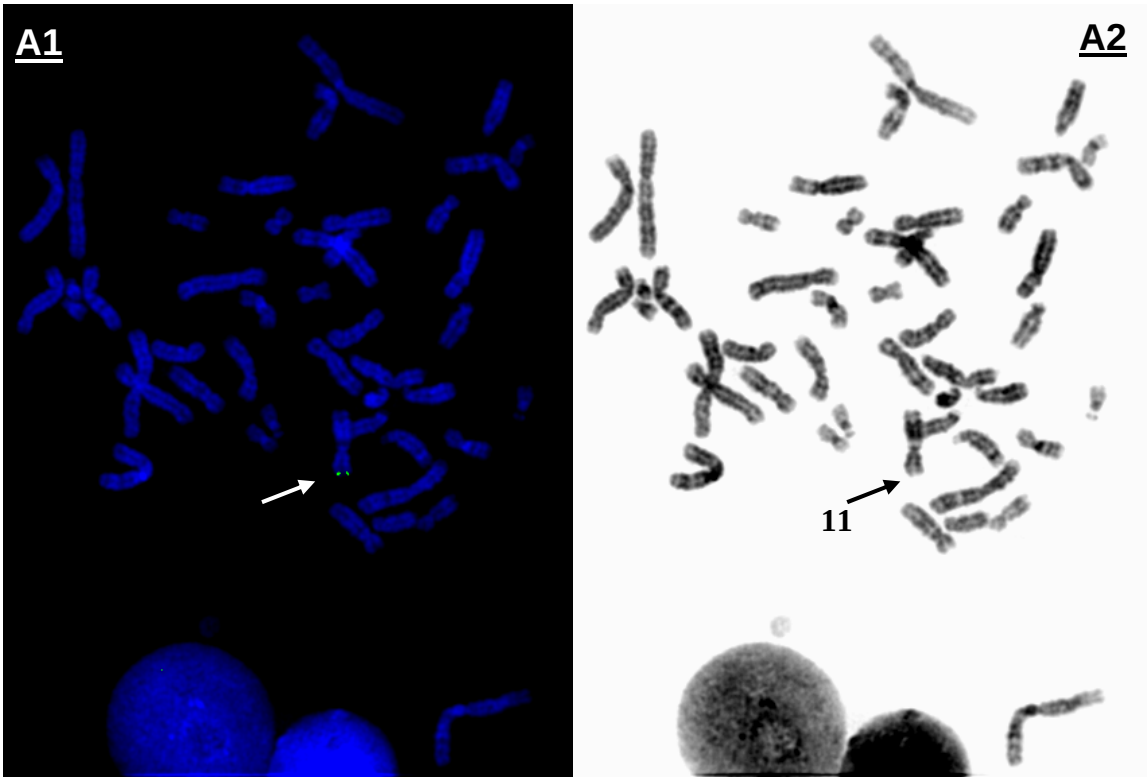


Bee Ling Ng

Figure 6.2 Flow karyograms for patients AMD, AS and KK. In all 3 patients, the integration of the virus had a negligible effect on the karyogram and therefore the chromosome carrying the integration was unable to be determined by this method.

6.2 FISH analysis of viral integration sites in 3 patients

A FISH probe created from viral genomic DNA was hybridised onto metaphase chromosome spreads isolated from the patients' cell-lines. Analysis of the signal revealed that in each patient the virus had integrated into one homologue of a particular chromosome only. In patient AMD the virus had integrated into 11p15.2-11p15.5, in patient AS it had integrated into 9q34.12-9q34.3 and in patient KK it had integrated into 17p13.1-17p13.3 (Figure 6.3). In all cases, the integration sites were seen to be within or close to the telomeres of human chromosomes consistent with published data where integration had been observed at 1q44, 22q13 and 17p13.3 (Daibata et al. 1998; Daibata et al. 1999; Morris et al. 1999).



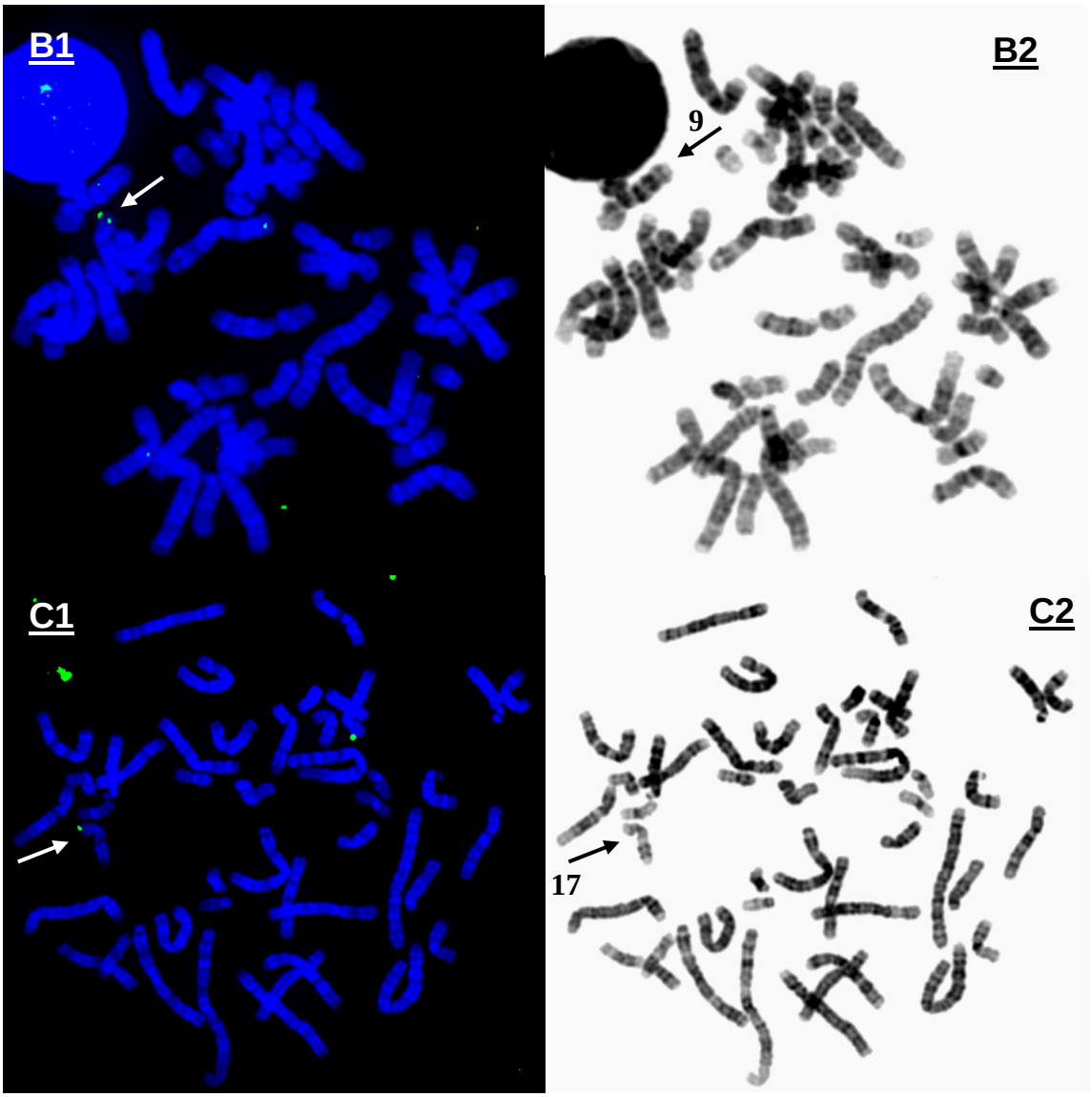


Figure 6.3 Metaphase chromosome analysis of viral integration in patients AMD, AS and KK. 1 metaphase FISH image, 2 inverted DAPI banding of metaphase chromosomes. Results for A patient AMD at 11p15.2-11p15.5, B patient AS at 9q34.12-9q34.3 and C patient KK at 17p13.1-17p13.3.

6.3 Investigation of integration using genomic microarrays

A custom-made genomic microarray was constructed for the quantification of HHV-6B integration into the human genome. The microarray contained a selection of plasmid clones covering 103,922bp of the genome as shown in Figure 6.1. In addition, to increase the resolution of the analysis, the microarray contained 158 PCR products designed to amplify approximately 800-1200bp regions of DNA to tile across the 144.5Kb of unique sequence of the HHV-6 genome. The minimum overlap between PCR products was 9bp and the maximum overlap was 1145bp.

6.3.1 Verification of the microarray

Prior to using the microarray for analysis of viral integration in the 3 patients, the clones on the array assessed for Cy3 versus Cy5 labelling bias using a self versus self hybridisation. The experiment was performed using DNA from patient AMD.

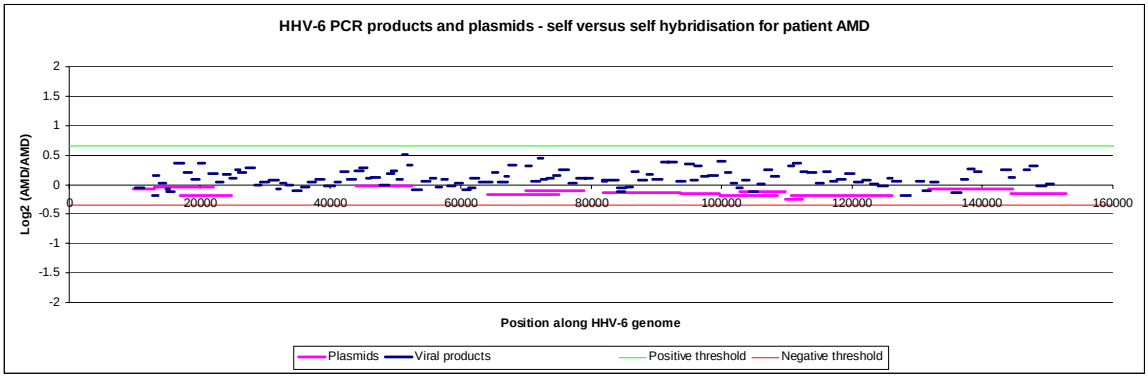


Figure 6.4 Self versus self hybridisation onto custom-made PCR product and plasmid microarray using patient AMD DNA.

All plasmid clones and PCR products were seen to lie within 2 standard deviations of the mean $\log_2$ ratio and therefore no elements on the microarray were excluded on the basis of this experiment.

6.3.2 Comparison between the 3 patients

In order to investigate any differences between the viral copy number integrated into each patient, all three patients were hybridised to the array against each other. Patient AMD was hybridised against patient AS, patient AS hybridised against patient KK and patient AMD against patient KK. Figure 6.5 shows the results of this comparison.

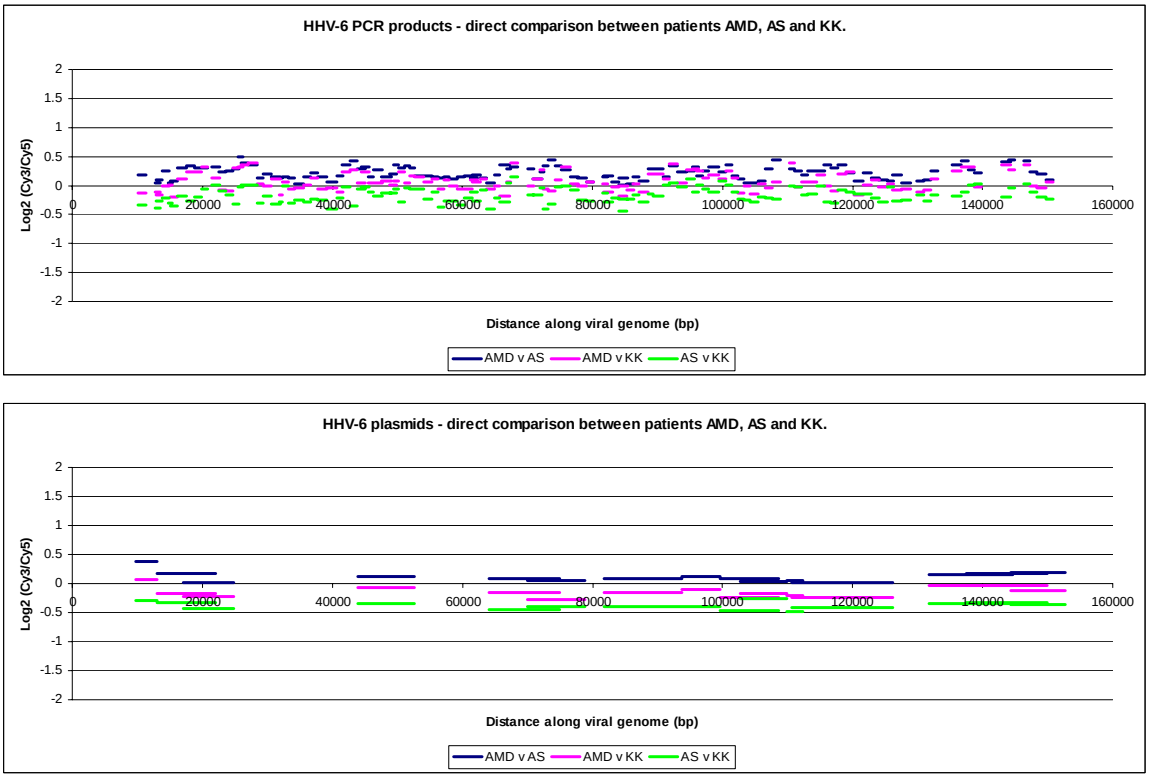


Figure 6.5 Comparative analysis of viral copy number in patients AMD, AS and KK on a custom-made array containing PCR products and plasmids.

Analysis of the average ratio across the PCR products for each hybridisation suggested that patient AMD had a greater number of viral genomes integrated than patient KK which had more than patient AS. The patient samples were analysed comparatively as it was not possible to determine the exact number of integrated viral genomes per patient because there was no sample with a known copy number.

6.4 Investigation of integration using a custom-made library

Patient KK was selected for further analysis because previous FISH analysis had shown that the virus had integrated into chromosome 17p13.1-17p13.3 (Figure 6.3). Patients AMD and AS had shown integration sites on chromosomes 11 and 9 respectively. Neither of these chromosomes can be flow sorted cleanly due to their similar size and constitution to chromosomes 10 and 12. A custom-made fosmid library was created using flow sorted DNA from patient KK to investigate the structure of the integrated virus.

6.4.1 Generation of custom-made library

Chromosomes containing the integrated viral genome along with their normal homologues were isolated from the remainder of the genome by flow sorting (Figure 6.6).

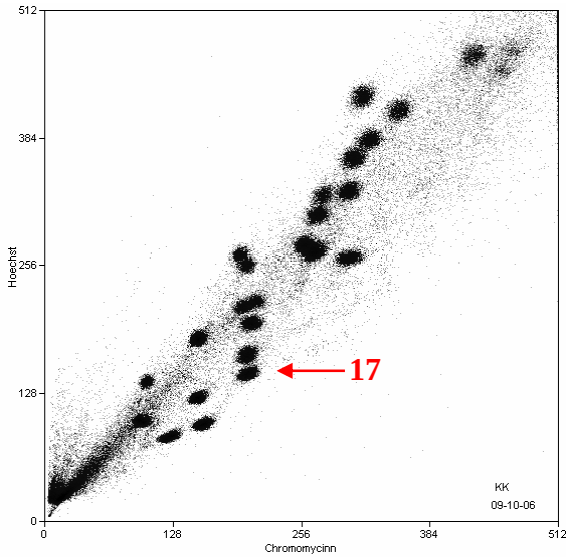


Figure 6.6 Flow karyogram for patient KK with integrated virus on 17p13.1-17p13.3. Flow sorting isolated chromosome 17 from the remainder of the genome, but only half of the flow sorted chromosomes contained integrated virus.

The flow sorted chromosomes were sheared by passing twice through a syringe needle to create an average fragment size of approximately 36Kb (Figure 6.7) and used to generate a custom-made fosmid library.

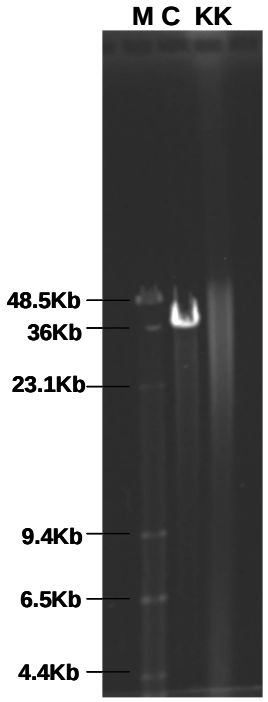


Figure 6.7 Results showing flow sorted chromosome 17 DNA after shearing. KK; sheared patient KK DNA run against M; λ DNA digested with HindIII and C; control DNA of 36Kb supplied with the library production kit.

6.4.2 Screening of custom-made library

After plating of the library, 18,432 colonies were picked into gridded plates by robots, resulting in a theoretical eightfold coverage along the length of chromosome 17. These gridded clones were used to generate filters for library screening (as described in the Methods).

PCR primer pairs were designed to amplify products every 10Kb along the central unique segment of the viral genome. 15 products (Appendix A6) were labelled and hybridised to the gridded library. Eight clones were identified as containing viral DNA; KK_1H19, KK_7B16, KK_8J11, KK_17G11, KK_23M7,

KK_23M8, KK_26O3 and KK_28P16. These clones were end sequenced to verify whether they contained viral DNA only or whether they were chimeric for human and viral DNA and therefore contained the sequence across the integration sites. The end sequence results are summarised in Table 6.1.

Clone name	Left end read	Right end read
KK_1H19	Human; 17;51,582,334-51,582,799	Human; 17;51,538,661-51,539,359
KK_7B16	Virus; 58,660-58,813	Virus; 23,275-23,983
KK_8J11	Virus; 85,739-86,404	Virus; 47,116-47,693
KK_17G11	Virus; 35,848-35,914	Virus; 73,313-73,685
KK_23M7	Virus; 60,127-60,705	Virus; 94,673-95,421
KK_23M8	None	None
KK_26O3	None	None
KK_28P16	None	None

Table 6.1 Alignment of end sequence data to human and viral genomes.

The end sequence reads for clone KK_1H19 aligned to chromosome 17q23.2 approximately 20Mb proximal to the integration site detected by FISH analysis. This data suggested that either the well contained a mixed clone or that whilst the PCR primers used to generate products for library screening were specific for the viral genome, the product created had sufficient homology to hybridise to chromosome 17q23.2.

Four clones; KK7_B16 (35,538bp), KK_8J11 (39,288bp), KK_17G11 (37,465bp) and KK_23M7 (35,294bp) aligned to viral DNA only (Figure 6.8). Analysis of the alignments revealed that the data was consistent with the viral genome integrating intact with no rearrangements.

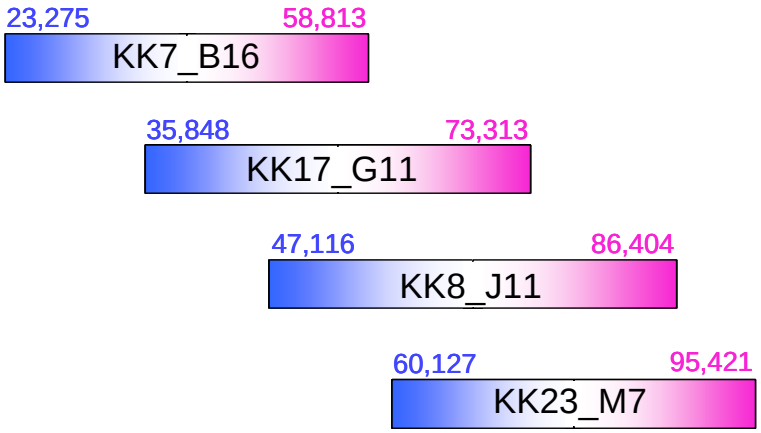


Figure 6.8 Relative positions of the 4 fosmid clones identified through screening of the patient KK library using PCR products spaced every 10Kb along the viral genome.

Three further clones failed to produce end sequence reads; KK_23M8, KK_26O3 and KK_28P16. These clones showed only weak hybridisation signals during the screen suggesting that these are not genuine positive clones.

No clones chimeric for viral and human chromosome 17 sequence were identified.

A second screen of the library was performed using primers designed approximately every 1Kb in the 10Kb of unique sequence close to the direct repeats of the viral genome (Appendix A7). A pool of 9 products was generated for the left end and a pool of 7 products for the right end of the viral genome (pictured in Figure 6.9 with the whole genome primers used in the original screen). However, these PCR product pools did not hybridise to any clones on the gridded filters. Due to time constraints, this experiment was not repeated to determine if this result was due to technical issues, or if the lack of hybridisation signal was indicative of the absence of fosmid clones containing sequence from the first and last 10Kb of the unique region of the viral genome.

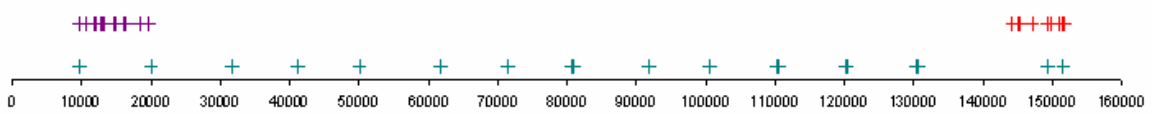


Figure 6.9 Summary of PCR products used for screening the patient KK custom-made fosmid library. The pool of 15 products used in the original screen are shown in green, the pool of 9 products at the left end of the genome are shown in purple and the pool of 7 products at the right end of the genome are shown in red along the length of the viral genome.

6.5 Investigation of integration using FISH onto extended DNA fibres

To verify whether the viral genome was integrated intact and with no rearrangement as hypothesised in section 6.4.2, probes were created from 2 plasmid clones and hybridised to extended chromatin fibres generated from patient KK (Figure 6.10). The plasmid clone from the left end of the viral genome was pHZV14, with an insert size of 8.7Kb and the plasmid clone from the right end of the genome was pH6Z-204 with an insert size of 12.3Kb.

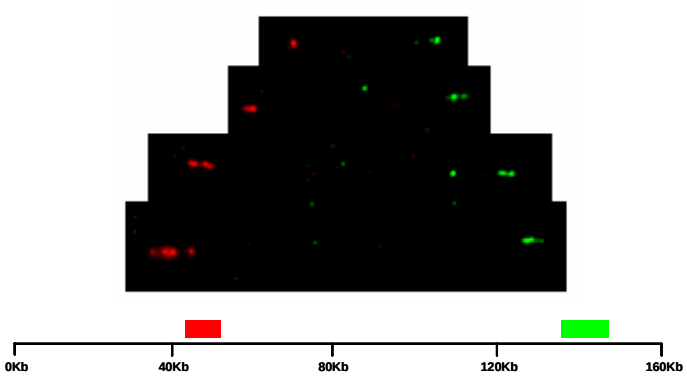


Figure 6.10 Results of plasmid clone FISH hybridisation onto extended chromatin fibres created from patient KK. pHZV14 was detected with Texas Red (red) and pH6Z-204 detected with FITC (green).

The FISH results showed that the 2 plasmids used as probes show a relationship with a consistent gap of approximately 15 times the size of the probes. This suggested that the integrated virus was complete with no rearrangement.

Occasional relationships where the 2 probes were seen to lie extremely close to each other were also observed, however, it was unclear whether this relationship was real suggesting that multiple copies of the virus integrated tandemly, or whether the signals were chance events resulting from the distribution of fibres along the slide.

6.6 Conclusions

The techniques used in this chapter were initially utilised to map rearrangement breakpoints in the human genome and have also been used to investigate viral integration. FISH onto metaphase chromosomes identified the sites of integration in 3 patients into or close to the telomeres of 9q, 11p and 17p and array CGH distinguished that there was variation in the number of integrated viral genomes between the 3 patients. The generation and screening of a custom-made fosmid library showed that the structure of the integrated virus appeared to be intact without rearrangement within the unique region of the genome. Further work to isolate the fosmid clones chimeric for human and viral DNA containing the integration sites would elucidate the sequence across the integration junction and confirm or refute the hypothesis that the HHV-6 virus integrates by way of its human telomeric like repeats.

7 Discussion

7.1 Methods of rearrangement breakpoint mapping

7.1.1 Breakpoint mapping using microarray based techniques

Apparently balanced translocations in 3 phenotypically abnormal patients were mapped using a variety of microarray based strategies prior to sequencing. The breakpoints were mapped by array painting onto a 1Mb microarray followed by FISH analysis of intermediate BAC clones prior to this study. The breakpoints were mapped further by array painting onto custom-made fosmid microarrays followed by STS PCR mapping prior to LR PCR for the amplification of junction fragments and subsequent sequencing. In the case of patient t(2;7)(q37.3;p15.1), STS PCR failed to refine the chromosome 7 translocation breakpoint further than array painting onto the custom-made fosmid microarray so alternative strategies of array painting onto custom-made PCR product microarrays and custom-made oligonucleotide microarrays with increased resolution were adopted.

The advancement of genomic microarray technology during the course of this project has increased the resolution from approximately 1Mb to under 50bp, increasing the speed and accuracy with which rearrangement breakpoints can be mapped. Our preferred strategy of translocation breakpoint mapping is currently array painting onto a Whole Genome Tile Path microarray followed by array painting onto a custom-made oligonucleotide microarray, thus enabling design of primers for amplification of the rearrangement junction fragments by PCR after just two hybridisations. In the future, the increasing resolution of microarrays may allow this to be reduced to a single hybridisation on one microarray. Patient t(2;7)(q37.3;p15.1) was used to directly compare these methodologies (Figure 7.1).

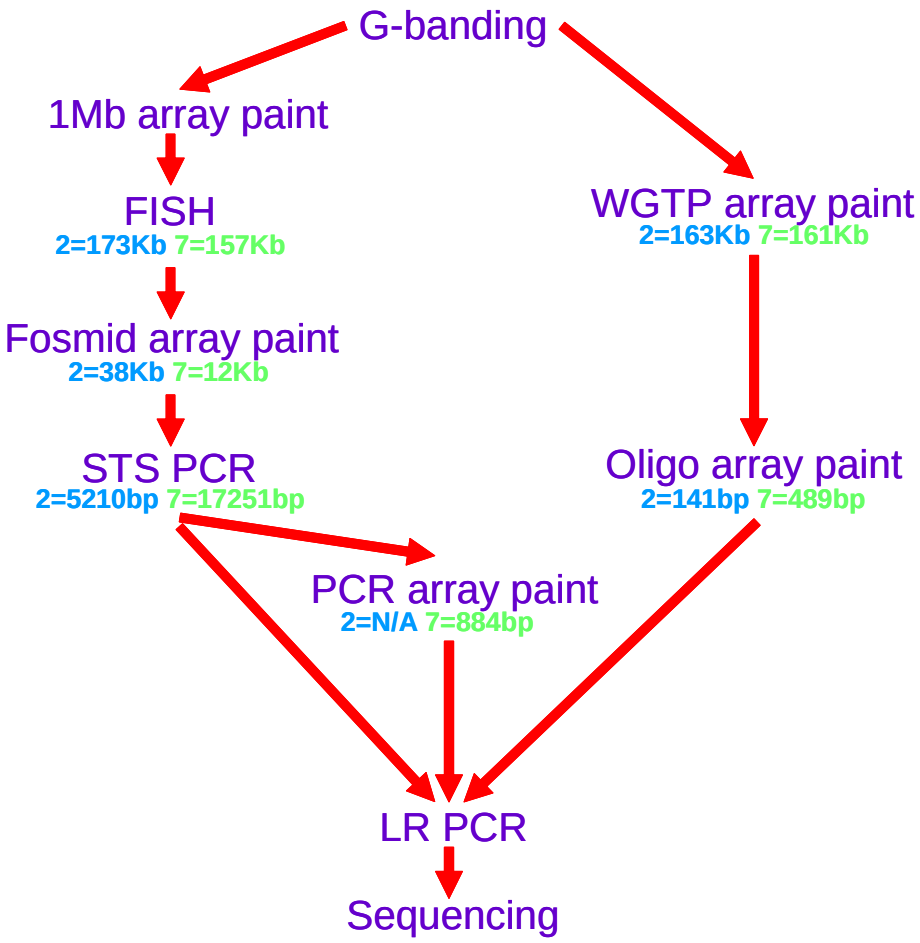


Figure 7.1 Direct comparison of strategies used for translocation breakpoint mapping using patient t(2;7;)(q37.3;p15.1) as a test case. The mapped intervals for the chromosome 2 breakpoint are shown in blue and the chromosome 7 breakpoint in green as defined by each technique.

The quantitative nature of array painting was demonstrated by further analysis of the array painting data generated by the custom-made fosmid microarrays. Interpolation of the data predicted that the chromosome 2 breakpoint fell at 236,556,071bp and chromosome 7 breakpoint fell at 30,984,039bp. This prediction was 7,492bp away from the actual breakpoint for chromosome 2 and 436bp away from the chromosome 7 breakpoint. This data emphasised the need for the generation of microarrays with a high redundancy of clones rather than

minimal tiling path coverage to enable the prediction of breakpoint positions from the ratios obtained from overlapping clones.

Full analysis of the chromosome 7 breakpoint region in patient t(2;7)(q37.3;p15.1) showed the complicated nature of the genome (Figure 7.2).

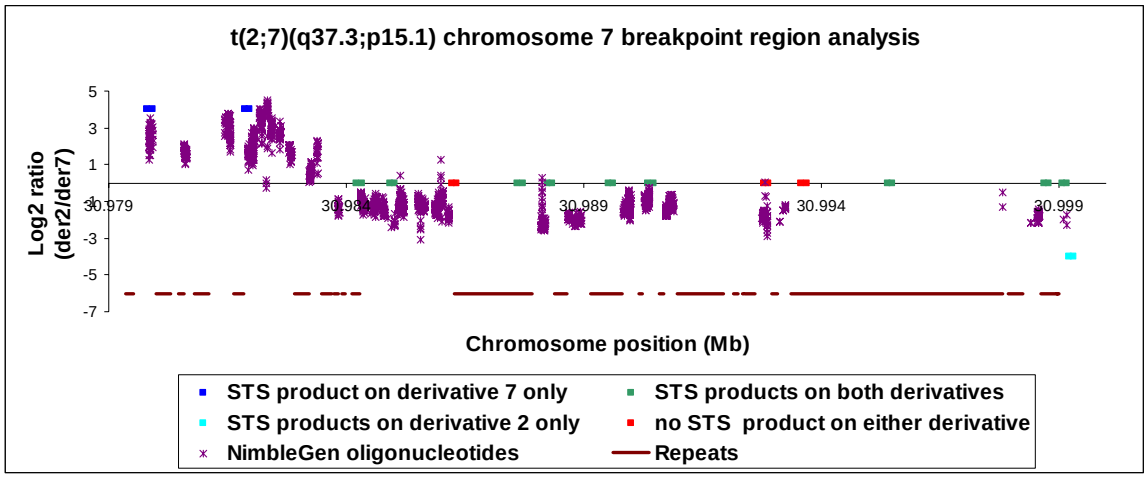


Figure 7.2 Analysis of STS PCR and oligonucleotide microarray array painting results for patient t(2;7)(q37.3;p15.1) with repeat content of surrounding region.

Only oligonucleotides considered to be unique within the human genome were selected for generation of the microarray resulting in incomplete coverage across the breakpoint region. As expected, closer inspection of the sequence revealed that gaps in the oligonucleotide coverage corresponded to known repeat elements annotated in the USCS web browser. STS PCR analysis across this region produced conflicting results due to the repetitive nature of the sequence. This complexity may have impeded whole genome amplification by DOP PCR performed prior to STS PCR mapping, resulting in incomplete amplification of the breakpoint region.

The techniques for translocation breakpoint mapping were also used to successfully map and sequence across a duplication junction in patient t(2;7)(q37.3;p15.1) with an additional duplication of 3p26.3. LR PCR primers were designed to amplify across a junction resulting from a simple tandem duplication. If these primers had failed to amplify a fragment, further investigation by FISH onto extended chromatin fibres using clones from the breakpoint regions as probes would have been required to resolve the orientation of the duplication prior to PCR amplification.

7.1.2 Translocation breakpoint mapping using custom-made fosmid libraries

The mapping of translocation breakpoints by array painting and PCR relies on the assumption that the rearrangement is simple. Translocations which are accompanied by cryptic imbalances or inversions around the breakpoints will be harder to map by these methods and in these instances, alternative strategies of breakpoint mapping would be required. For this reason we investigated the generation and screening of a custom-made library for breakpoint mapping as summarised in Figure 7.3. Custom-made libraries have previously been used to identify aberrations in a breast cancer cell-line (Volik et al. 2003). A BAC library created with 0.37 fold coverage was end sequenced and the data mapped back onto the human genome reference sequence to identify rearrangements within the cell line.

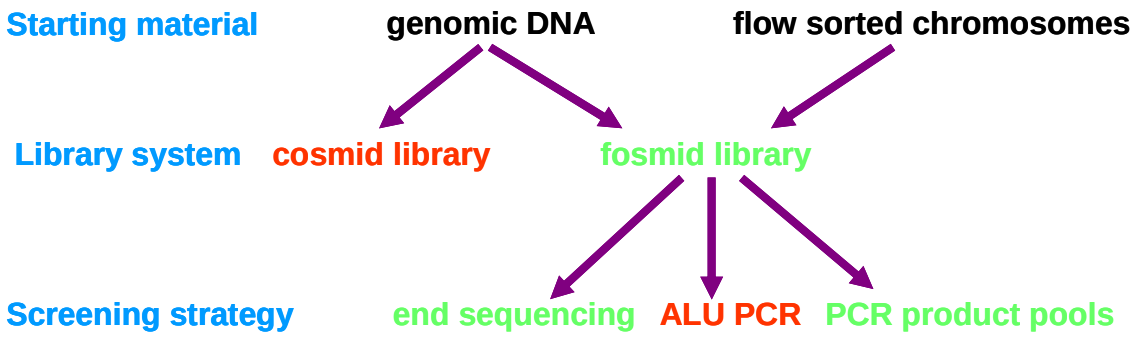


Figure 7.3 Schematic summarising the approaches used to generate and screen a custom-made library. Techniques in green were successful.

Using genomic DNA extracted from a patient derived cell-line, 2 protocols for the production of a custom-made library were investigated concurrently; a cosmid library and a fosmid library. The fosmid library proved to be a successful method, but the cosmid library failed to produce any colonies. The cosmid library technique was not investigated further but the lack of cosmid clones may have been due to technical issues such as loss of template material and reagents that were not optimal. Subsequently, flow sorted derivative chromosomes were used as starting material for the fosmid library in order to increase the purity and coverage of the library created.

A fosmid library was created from flow sorted derivative chromosomes from patient t(7;13)(q31.3;q21.3) to test the feasibility of this approach. All clones from the library were end sequenced and screened using radiolabelled Alu PCR products and a pool of radiolabelled PCR products. Alu PCR product screening did not successfully identify any translocation spanning clones in the test libraries. This may have been due to the lack of amplification of human derived DNA using the primers specified. Whilst bands in the range of 100 to 1000bp were detected by gel electrophoresis after temperature cycling, *in silico* PCR analysis using the human genome reference sequence failed to predict any amplification under 3Kb (anything larger than 2Kb would be outside the limits of the PCR cycling conditions). PCR product pool screening of the fosmid library identified spanning clones in both libraries. End sequence screening identified a spanning clone in the derivative chromosome 7 library and mapped the breakpoint to a 130Kb region between clones in the derivative chromosome 13 library. In addition, the end sequence data obtained provided further information about the coverage of the library along the derivative chromosomes. The 2 successful strategies of screening by end sequencing and by radiolabelled PCR product pools have advantages and disadvantages which are detailed in Table 7.1.

	Advantage	Disadvantage
End sequencing	Automated process	Cost
	Additional data generated	
PCR product pool	Quick results (4 days)	Radiation hazards

Table 7.1 Summary of major advantages to screening by end sequencing or by radiolabelled PCR product pools.

Analysis of the full sequence data generated from the chimeric clones showed the breakpoint positions were consistent with those mapped by array painting and PCR.

After the assessment of production and screening methods using patient t(7;13)(q31.3;q21.3) as a test case, a library was generated from a patient with a previously uncharacterised t(2;6)(q21.1;q25.1) translocation. The translocation breakpoints were initially mapped by array painting flow sorted derivative chromosomes onto a whole genome tile path microarray to identify spanning BAC clones. The breakpoint spanning fosmid clones from the custom-made library were successfully identified using radiolabelled PCR product pools generated from the spanning BAC clone information. The fosmid clones were isolated and sequenced allowing the identification of the precise breakpoint positions and confirming the simple nature of the rearrangement. It was noted that 7bp were deleted from chromosome 2 and 3bp from chromosome 6. Analysis of the extra sequence data (approximately 40Kb for each clone) shared 100% homology with the human genome reference sequence.

7.1.3 Application of techniques developed for the mapping of rearrangement breakpoints to the investigation of viral integration in the human genome

One of the aims of this thesis was to demonstrate that the techniques developed for the mapping of translocation breakpoints could also be applied to other rearrangements such as viral integration. A custom-made fosmid library was generated using flow sorted chromosomes from a patient carrying integrated

HHV-6 DNA. Whilst the generation and screening of the library has not, so far, identified clones chimeric for human and viral DNA and therefore spanning the breakpoint, further screening with alternative probes may do so. For example, using a human telomere probe in conjunction with probes generated towards the ends of the viral genome should identify chimeric clones as it is possible that the telomeric like repeat present in the virus enables the virus to integrate into the human telomeres by homologous recombination. A library with a greater depth of coverage along the chromosome and a pool of viral probes with more comprehensive tiling along the HHV-6 genome could also be investigated.

In addition, conventional techniques could also be applied to resolve the integration sites; Southern blots of the chromosome containing the integrated virus could be screened using human telomere and viral probes to isolate a chimeric fragment. Once a clone or fragment has been identified, full sequencing would generate data across the integration site. However, if as suspected, the integration occurs in the human telomere ($(TTAGGG)_n$) due to the telomeric like repeat present in the virus ($(GGGTTA)_n$) then it may be impossible to determine the exact integration site as long stretches of these repeats appear identical.

7.2 Bioinformatic analysis of genomic rearrangement breakpoints

Chromosome rearrangements have been associated with abnormal phenotypes; either by direct disruption of gene structures, modification of gene regulation or alteration of the copy number. Bioinformatic analysis of the regions surrounding the rearrangements may identify genes that may be causal to a phenotype in addition to genomic structures or motifs which may help to elucidate a mechanism behind the rearrangement.

7.2.1 Genotype to phenotype correlations

7.2.1.1 Identification of non-pathological variation

Analysis of human genomes is being conducted with increasing resolution. With this increase in resolution, greater complexity is found within the genome, but it is important to determine whether the variation seen between individuals is normal or pathological.

Genome-wide studies have identified large regions of variability within the human genome (Iafrate et al. 2004; Sebat et al. 2004; Tuzun et al. 2005). A recent study estimated that approximately 12% of the human genome was subject to copy number variation (CNV) (Redon et al. 2006). These regions of variation were identified in studies using apparently clinically normal individuals and are therefore assumed to be non-pathogenic. However, it should be noted that patients were described as clinically normal at the time of assessment and this assessment does not allow for extremely mild pathogenic phenotypes, or pathology that may develop later in life. All regions of variation identified by these studies have been summarised in the Database for Genomic Variants (http://projects.tcag.ca/variation/). In addition to large structural variation, single nucleotide polymorphisms (SNPs) are believed to be responsible for phenotypic variation. An investigation into the significance of SNPs and CNVs using expression level analyses of approximately 15,000 transcripts revealed that SNPs were responsible for 83.6% and CNVs 17.7% of the genetic variation detected in gene expression studied with little overlap (1.3%) (Stranger et al. 2007).

CNVs have dosage implications for the genes that they affect. A dosage sensitive gene within a region that is duplicated may lead to overexpression and a dosage sensitive gene in a region that is deleted may result in haploinsufficiency for that gene. For example, Charcot Marie Tooth disease (CMT1A) and hereditary

neuropathy with liability to pressure palsies (HNPP) are believed to be the result of overexpression and haploinsufficiency of the *PMP22* gene in duplications and deletions of chromosome 17p11.2 respectively (Patel et al. 1992; Chance et al. 1993). In order to ascertain whether a CNV is pathogenic, it is important to determine whether the variation is *de novo* or inherited.

Whole genome screening of patient t(2;7)(q37.3;p15.1) on a 1Mb resolution microarray revealed a duplication at 3p26.3. Subsequent CGH analysis on a custom-made fosmid microarray, followed by a custom-made oligonucleotide microarray refined the breakpoints sufficiently to allow direct amplification of the junction fragment. Sequencing of the junction fragment located the distal breakpoint within a 260bp region from 1,756,993 to 1,757,252bp and the proximal breakpoint within a 258bp region from 3,614,132 to 3,614,389bp. Analysis of the parental DNAs revealed that the same duplication was present in the phenotypically normal father, but not in the mother. Interrogation of the CNV data available in Ensembl revealed a known 289Kb CNV locus from 2,181,272 to 2,470,417bp. It was therefore unlikely that the duplication was pathogenic as the impact of an increase in copy number of genes within the duplicated region did not appear to be significant.

7.2.1.2 Gene disruption

Investigation into annotated genes at the breakpoint regions of the 3 patients studied revealed that 3 of the 6 translocation breakpoints directly disrupted a gene, providing candidate genes for 2 of the 3 patient phenotypes. Verification of the ensuing altered protein could be obtained through Western blotting and confirmation that the translocation is directly responsible for the phenotype could be achieved using expression studies to determine whether the truncated proteins are expressed. Murine knockout models could also be generated for comparative phenotype analysis.

In addition to chromosome rearrangements, mutations may have an effect on a patient's phenotype. In a case where a gene is not dosage sensitive, a translocation may not have an effect unless the other allele is affected in some way, for example, a simple mutation. In these cases, further analysis of the genes in the breakpoint regions would be required, for example resequencing of the gene exons for mutation detection.

It has been noted that translocation breakpoints in phenotypically normal patients are generally not accompanied by additional imbalance in direct contrast to the breakpoints in phenotypically abnormal patients which are often seen to be accompanied by cryptic imbalances (Baptista et al. 2005).

The risk of an apparently balanced reciprocal translocation having an associated congenital abnormality was estimated at 6.1% (Warburton 1991). However, the risk of a complex balanced rearrangement having an associated phenotype has been estimated at 23% (14 out of 60 cases studied) (Madan et al. 1997), indicating that the risk of an associated congenital abnormality rises with increasing complexity of the rearrangement.

The genotype to phenotype correlation for patient t(3;11)(q21;q12) is particularly complex. The patient is one of monochorionic, monoamniotic twins who by G-banding analysis carry the same translocation. Patient t(3;11)(q21;q12) presented with a congenital duodenal obstruction, complex congenital heart disease and facial dysmorphism whilst the sibling was clinically normal. As the twins are identical, it is unlikely that the translocation is responsible for the phenotype. The phenotypic difference between the twins could be due to discordant imprinting. Cases have been recorded where genetically identical twins present with different phenotypes as observed in Beckwith-Wiedermann syndrome (BWS). A study of 5 identical twins with discordant phenotypes

identified differences in the imprinting of the *KCNQ1OT1* gene as responsible for the variations observed in clinical diagnoses (Weksberg et al. 2002).

An alternative explanation for the clinical differences seen between patient t(3;11)(q21;q12) and her twin is X inactivation. Skewed X inactivation has been shown to be the cause of discordant phenotypes in a pair of twins where one twin presented with Duchenne Muscular Dystrophy and the other was clinically normal (Richards et al. 1990).

7.2.2 Mechanisms underlying genomic rearrangements

The sequence across the translocation junctions in all 3 patients investigated in Chapter 3 was characteristic of a mechanism of non-homologous end joining as there was a lack of homologous sequence between both the breakpoint regions and small numbers of bases were deleted and/or duplicated and/or inserted in all 6 of the breakpoints sequenced. In addition, analysis of the sequence surrounding the breakpoints did not identify any repeat structures or sequence motifs that were common to all the breakpoints. In patient t(2;7)(q37.3;p15.1) where a 19bp stretch of bases was inserted at the breakpoint, analysis of the sequence found no homology to the human genome reference sequence. A 41bp insertion of mitochondrial DNA had previously been identified at the derivative 9 junction in a patient with a t(9;11)(p24;q23) (Willett-Brozick et al. 2001) however, alignment of the 19bp insertion sequence in patient t(2;7)(q37.3;p15.1) to human mitochondrial DNA (Accession number NC_001807) failed to find any homology.

The mechanism behind the duplication seen in patient t(2;7)(q37.3;p15.1) was, in direct contrast, believed to be homologous recombination due to the Alu repeats found at both breakpoints. In fact, the precise duplication breakpoints could not be determined owing to the 97% homology between these repeats.

7.2.2.1 Predisposition

It is believed that certain genomic architecture can lead to a predisposition for disease. For example, in Williams Beuren Syndrome, 90% of patients have a 1.5Mb deletion at 7q11.23 where the breakpoints fall within low copy repeats (Nickerson et al. 1995; Bayes et al. 2003). Further analysis of the heredity of the disease revealed that 5% of the general population carry an inversion of the region between these repeat segments and in the next generation 33% of the offspring carry the deletion associated with Williams Beuren syndrome, suggesting that the inversion predisposes for the deletion (Osborne et al. 2001). This phenomenon is also believed to be true for Sotos syndrome; In a study of Sotos syndrome within the Japanese population, it was found that 14/14 fathers and 8/9 mothers of children carrying the associated 1.9Mb microdeletion carried a heterozygous inversion of the interval between the LCRs responsible in the paternally and maternally derived chromosomes respectively (Visser et al. 2005). More recently, a 900Kb inversion has been seen to predispose for a 500-650Kb deletion encompassing the *MAPT* locus at 17q21.31 in 3 patients with mental retardation and dysmorphic features (Shaw-Smith et al. 2006).

7.2.2.2 Occurrence of rearrangements

There are approximately 30-50,000 recombination hotspots representing 3% of the human genome equating to 1 every 50-100Kb. At least 80% of all recombination events occur in these short regions. A positive relationship between GC content and recombination has been noted with recombination rates found to be lower within genes (McVean et al. 2004).

Within the nucleus, chromosomes exist in discrete compartments known as chromosome territories where the most gene rich chromosomes (chromosomes 17,19 and 22) congregate at the centre of the nucleus whereas the more gene poor chromosomes gravitate towards the nuclear periphery (Boyle et al. 2001). A study of more than 10,000 constitutional translocations revealed that the larger

chromosomes are generally more involved in translocations, perhaps due to the length of DNA available for rearrangement (Bickmore and Teague 2002). It was also noted that the most gene dense chromosomes are less likely to be involved in translocations.

These studies may suggest that the human genome protects itself against rearrangements, however it must be considered that most studies have been conducted using patients rather than aborted conceptions so that the study group might be biased and that the majority of rearrangements that disrupt genes are lethal.

7.3 Recent developments and the future of breakpoint mapping

An alternative to the use of flow sorted chromosomes for array painting is the use of microdissected chromosomes for hybridisation to the microarrays. One study of 5 patients with apparently balanced reciprocal translocations used 4-6 microdissected derivative chromosomes for DOP PCR amplification prior to hybridisation to a 1 Mb resolution microarray (Backx et al. 2007). This procedure successfully mapped all 10 translocation breakpoints to 1Mb intervals.

International public efforts to sequence the human genome cost approximately $3 billion but the race is now on to sequence an individual's genome for $1000 (Service 2006). New sequencing technologies are being developed and the cost of sequencing is decreasing. One such development is the generation of sequence using 454 pyrosequencers (www.454.com). These machines fragment the genome into 300-500bp segments which are attached to beads, denatured and amplified in the presence of luciferase which fluoresces every time a nucleotide is added to the chain. By analysing the light flashes and the corresponding nucleotide present on each bead, the systematic growth and subsequent sequence can be tracked electronically. This 454 technology has been used to conduct a study of structural variation within the genomes of 2

individuals (Korbel et al. In press). The first was the female used to create a fosmid library to verify the human genome reference sequence as part of the Human Genome Project (IHGSC 2004) and more recently used to originally assess copy number variation by end sequencing (Tuzun et al. 2005). The second individual was a member of the group used to detect CNVs by array CGH (Redon et al. 2006). Genomic DNA from both individuals was fragmented to 3Kb and the paired ends of the fragments (equating to 2.1x and 4.3x coverage of the genomes) were sequenced and aligned to a reference genome to assess the level of structural variation. This technology enabled a 3Kb resolution scan of the genome to be performed with an average resolution of breakpoint assignment of 644bp (Korbel et al. 2007).

A more recent advancement of this method is to array the fragmented DNA directly onto microarray slides where they can be amplified *in situ* (www.illumina.com). Addition of all 4 fluorescently labelled nucleotides allow each cycle of the sequencing reaction to occur simultaneously, and the fluorescence of each incorporated base to be detected. This technology is capable of sequencing across long stretches of homopolymers.

Analysis of the read depths obtained using these sequencing technologies will provide data on any copy number variation within a genome. For example, a region showing amplification within a genome will exhibit a higher read depth than at a region with a normal copy number. These technologies could potentially be used to sequence derivative chromosomes in patients carrying rearrangements; direct comparison of the data generated with the human genome reference sequence would identify any differences between the 2 genomes including translocations, inversion and copy number differences. These strategies of breakpoint mapping would be of particular use in cases with complex rearrangements.

7.4 Conclusions

The advancement of chromosome analysis has improved the speed and resolution with which chromosome rearrangements can be studied. By increasing the efficiencies of methodologies for the mapping and subsequent sequencing of rearrangement breakpoints the number of published sequenced constitutional translocation breakpoints will be expanded, so that a comprehensive study of the breakpoint regions may help to elucidate the mechanism causing the rearrangement. The data so far suggests that the mechanism by which these non-recurrent rearrangements occurs is non-homologous end joining, in direct contrast to recurrent rearrangements which are mediated by homologous recombination.

This thesis has developed strategies for the rapid analysis of rearrangement breakpoints and viral integration sites in the human genome. New, higher resolution microarrays have been applied to the technique of translocation breakpoint mapping and the generation and screening of custom-made fosmid libraries have been successfully used to identify the breakpoints in a balanced reciprocal translocation. The technique of breakpoint mapping using custom-made libraries enables the mapping of more complex rearrangements where analysis by microarrays and PCR might fail. This thesis also demonstrates that these techniques are applicable to the investigation of other genomic rearrangement studies such as viral integration into the human genome.

Aberle SW, Mandl CW, Kunz C, Popow-Kraupp T (1996) Presence of human herpesvirus 6 variants A and B in saliva and peripheral blood mononuclear cells of healthy adults. J Clin Microbiol 34:3223-3225
Abeysinghe SS, Chuzhanova N, Krawczak M, Ball EV, Cooper DN (2003) Translocation and gross deletion breakpoints in human inherited disease and cancer I: Nucleotide composition and recombination-associated motifs. Hum Mutat 22:229-244
Aerts S, Lambrechts D, Maity S, Van Loo P, Coessens B, De Smet F, Tranchevent LC, De Moor B, Marynen P, Hassan B, Carmeliet P, Moreau Y (2006) Gene prioritization through genomic data fusion. Nat Biotechnol 24:537-544
Albertsen HM, Abderrahim H, Cann HM, Dausset J, Le Paslier D, Cohen D (1990) Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. Proc Natl Acad Sci U S A 87:4256-4260
Aldred MA, Sanford RO, Thomas NS, Barrow MA, Wilson LC, Brueton LA, Bonaglia MC, Hennekam RC, Eng C, Dennis NR, Trembath RC (2004) Molecular analysis of 20 patients with 2q37.3 monosomy: definition of minimum deletion intervals for key phenotypes. J Med Genet 41:433-439
Aoki K, Suzuki K, Sugano T, Tasaka T, Nakahara K, Kuge O, Omori A, Kasai M (1995) A novel gene, Translin, encodes a recombination hotspot binding protein associated with chromosomal translocations. Nat Genet 10:167-174
Arai E, Ikeuchi T, Nakamura Y (1994) Characterization of the translocation breakpoint on chromosome 22q12.2 in a patient with neurofibromatosis type 2 (NF2). Hum Mol Genet 3:937-939
Astbury C, Christ LA, Aughton DJ, Cassidy SB, Kumar A, Eichler EE, Schwartz S (2004) Detection of deletions in de novo "balanced" chromosome rearrangements: further evidence for their role in phenotypic abnormalities. Genet Med 6:81-89
Backx L, Van Esch H, Melotte C, Kosyakova N, Starke H, Frijns JP, Liehr T, Vermeesch JR (2007) Array painting using microdissected chromosomes to map chromosomal breakpoints. Cytogenet Genome Res 116:158-166
Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE (2002) Recent segmental duplications in the human genome. Science 297:1003-1007
Baptista J, Prigmore E, Gribble SM, Jacobs PA, Carter NP, Crolla JA (2005) Molecular cytogenetic analyses of breakpoints in apparently balanced reciprocal translocations carried by phenotypically normal individuals. Eur J Hum Genet
Bauman JG, Wiegant J, Borst P, van Duijn P (1980) A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA. Exp Cell Res 128:485-490
Bayes M, Magano LF, Rivera N, Flores R, Perez Jurado LA (2003) Mutational mechanisms of Williams-Beuren syndrome deletions. Am J Hum Genet 73:131-151

Been MD, Burgess RR, Champoux JJ (1984) Nucleotide sequence preference at rat liver and wheat germ type 1 DNA topoisomerase breakage sites in duplex SV40 DNA. Nucleic Acids Res 12:3097-3114

Bhalla K, Phillips HA, Crawford J, McKenzie OL, Mulley JC, Eyre H, Gardner AE, Kremmidiotis G, Callen DF (2004) The de novo chromosome 16 translocations of two patients with abnormal phenotypes (mental retardation and epilepsy) disrupt the A2BP1 gene. J Hum Genet 49:308-311

Bickmore WA, Teague P (2002) Influences of chromosome size, gene density and nuclear position on the frequency of constitutional translocations in the human population. Chromosome Res 10:707-715

Bien-Willner GA, Stankiewicz P, Lupski JR (2007) SOX9cre1, a cis-acting regulatory element located 1.1 Mb upstream of SOX9, mediates its enhancement through the SHH pathway. Hum Mol Genet 16:1143-1156

Bocciardi R, Giorda R, Marigo V, Zordan P, Montanaro D, Gimelli S, Seri M, Lerone M, Ravazzolo R, Gimelli G (2005) Molecular characterization of a t(2;6) balanced translocation that is associated with a complex phenotype and leads to truncation of the TCBA1 gene. Hum Mutat 26:426-436

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

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

Boehm T, Mengle-Gaw L, Kees UR, Spurr N, Lavenir I, Forster A, Rabbitts TH (1989) Alternating purine-pyrimidine tracts may promote chromosomal translocations seen in a variety of human lymphoid tumours. Embo J 8:2621-2631

Bonaglia MC, Giorda R, Borgatti R, Felisari G, Gagliardi C, Selicorni A, Zuffardi O (2001) Disruption of the ProSAP2 gene in a t(12;22)(q24.1;q13.3) is associated with the 22q13.3 deletion syndrome. Am J Hum Genet 69:261-268

Bonora E, Lamb JA, Barnby G, Sykes N, Moberly T, Beyer KS, Klauck SM, et al. (2005) Mutation screening and association analysis of six candidate genes for autism on chromosome 7q. Eur J Hum Genet 13:198-207

Borg I, Freude K, Kubart S, Hoffmann K, Menzel C, Laccone F, Firth H, Ferguson-Smith MA, Tommerup N, Ropers HH, Sargan D, Kalscheuer VM (2005) Disruption of Netrin G1 by a balanced chromosome translocation in a girl with Rett syndrome. Eur J Hum Genet 13:921-927

Borg I, Squire M, Menzel C, Stout K, Morgan D, Willatt L, O'Brien PC, Ferguson-Smith MA, Ropers HH, Tommerup N, Kalscheuer VM, Sargan DR (2002) A cryptic deletion of 2q35 including part of the PAX3 gene detected by breakpoint mapping in a child with autism and a de novo 2;8 translocation. J Med Genet 39:391-399

Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA, Bickmore WA (2001) The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. Hum Mol Genet 10:211-219

Briggs M, Fox J, Tedder RS (1988) Age prevalence of antibody to human herpesvirus 6. Lancet 1:1058-1059
Budarf ML, Collins J, Gong W, Roe B, Wang Z, Bailey LC, Sellinger B, Michaud D, Driscoll DA, Emanuel BS (1995) Cloning a balanced translocation associated with DiGeorge syndrome and identification of a disrupted candidate gene. Nat Genet 10:269-278
Cargile CB, Goh DL, Goodman BK, Chen XN, Korenberg JR, Semenza GL, Thomas GH (2002) Molecular cytogenetic characterization of a subtle interstitial del(3)(p25.3p26.2) in a patient with deletion 3p syndrome. Am J Med Genet 109:133-138
Carvalho B, Ouwerkerk E, Meijer GA, Ylstra B (2004) High resolution microarray comparative genomic hybridisation analysis using spotted oligonucleotides. J Clin Pathol 57:644-646
Caspersson T, Farber S, Foley GE, Kudynowski J, Modest EJ, Simonsson E, Wagh U, Zech L (1968) Chemical differentiation along metaphase chromosomes. Exp Cell Res 49:219-222
Caspersson T, Lomakka G, Zech L (1972) The 24 fluorescence patterns of the human metaphase chromosomes - distinguishing characters and variability. Hereditas 67:89-102
Chance PF, Alderson MK, Leppig KA, Lensch MW, Matsunami N, Smith B, Swanson PD, Odelberg SJ, Disteche CM, Bird TD (1993) DNA deletion associated with hereditary neuropathy with liability to pressure palsies. Cell 72:143-151
Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SP (1996) Accessing genetic information with high-density DNA arrays. Science 274:610-614
Chen SJ, Chen Z, d'Auriol L, Le Coniat M, Grausz D, Berger R (1989) Ph1+bcr- acute leukemias: implication of Alu sequences in a chromosomal translocation occurring in the new cluster region within the BCR gene. Oncogene 4:195-202
Cheung AK, Hoggan MD, Hauswirth WW, Berns KI (1980) Integration of the adeno-associated virus genome into cellular DNA in latently infected human Detroit 6 cells. J Virol 33:739-748
Cheung VG, Nowak N, Jang W, Kirsch IR, Zhao S, Chen XN, Furey TS, et al. (2001) Integration of cytogenetic landmarks into the draft sequence of the human genome. Nature 409:953-958
Ciccone R, Giorda R, Gregato G, Guerrini R, Giglio S, Carrozzo R, Bonaglia MC, Priolo E, Lagana C, Tenconi R, Rocchi M, Pramparo T, Zuffardi O, Rossi E (2005) Reciprocal translocations: a trap for cytogenetists? Hum Genet 117:571-582
Clark DA, Ait-Khaled M, Wheeler AC, Kidd IM, McLaughlin JE, Johnson MA, Griffiths PD, Emery VC (1996) Quantification of human herpesvirus 6 in immunocompetent persons and post-mortem tissues from AIDS patients by PCR. J Gen Virol 77 (Pt 9):2271-2275
Cox JJ, Holden ST, Dee S, Burbridge JI, Raymond FL (2003) Identification of a 650 kb duplication at the X chromosome breakpoint in a patient with

46,X,t(X;8)(q28;q12) and non-syndromic mental retardation. J Med Genet 40:169-174
Crawford DC, Bhangale T, Li N, Hellenthal G, Rieder MJ, Nickerson DA, Stephens M (2004) Evidence for substantial fine-scale variation in recombination rates across the human genome. Nat Genet 36:700-706
Crisponi L, Uda M, Deiana M, Loi A, Nagaraja R, Chiappe F, Schlessinger D, Cao A, Pilia G (2004) FOXL2 inactivation by a translocation 171 kb away: analysis of 500 kb of chromosome 3 for candidate long-range regulatory sequences. Genomics 83:757-764
Daibata M, Taguchi T, Nemoto Y, Taguchi H, Miyoshi I (1999) Inheritance of chromosomally integrated human herpesvirus 6 DNA. Blood 94:1545-1549
Daibata M, Taguchi T, Sawada T, Taguchi H, Miyoshi I (1998) Chromosomal transmission of human herpesvirus 6 DNA in acute lymphoblastic leukaemia. Lancet 352:543-544
David D, Cardoso J, Marques B, Marques R, Silva ED, Santos H, Boavida MG (2003) Molecular characterization of a familial translocation implicates disruption of HDAC9 and possible position effect on TGFbeta2 in the pathogenesis of Peters' anomaly. Genomics 81:489-503
Demir E, Sabatelli P, Allamand V, Ferreiro A, Moghadaszadeh B, Makrelouf M, Topaloglu H, Echenne B, Merlini L, Guicheney P (2002) Mutations in COL6A3 cause severe and mild phenotypes of Ullrich congenital muscular dystrophy. Am J Hum Genet 70:1446-1458
Dhami P, Coffey AJ, Abbs S, Vermeesch JR, Dumanski JP, Woodward KJ, Andrews RM, Langford C, Vetrie D (2005) Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome. Am J Hum Genet 76:750-762
Downing RG, Sewankambo N, Serwadda D, Honess R, Crawford D, Jarrett R, Griffin BE (1987) Isolation of human lymphotropic herpesviruses from Uganda. Lancet 2:390
Duba HC, Doll A, Neyer M, Erdel M, Mann C, Hammerer I, Utermann G, Grzeschik KH (2002) The elastin gene is disrupted in a family with a balanced translocation t(7;16)(q11.23;q13) associated with a variable expression of the Williams-Beuren syndrome. Eur J Hum Genet 10:351-361
Edelmann L, Spiteri E, Koren K, Pulijaal V, Bialer MG, Shanske A, Goldberg R, Morrow BE (2001) AT-rich palindromes mediate the constitutional t(11;22) translocation. Am J Hum Genet 68:1-13
Edelmann L, Spiteri E, McCain N, Goldberg R, Pandita RK, Duong S, Fox J, Blumenthal D, Lalani SR, Shaffer LG, Morrow BE (1999) A common breakpoint on 11q23 in carriers of the constitutional t(11;22) translocation. Am J Hum Genet 65:1608-1616
Fantes J, Redeker B, Breen M, Boyle S, Brown J, Fletcher J, Jones S, Bickmore W, Fukushima Y, Mannens M, et al. (1995) Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. Hum Mol Genet 4:415-422

Fernandez T, Morgan T, Davis N, Klin A, Morris A, Farhi A, Lifton RP, State MW (2004) Disruption of contactin 4 (CNTN4) results in developmental delay and other features of 3p deletion syndrome. Am J Hum Genet 74:1286-1293

Fiegler H, Carr P, Douglas EJ, Burford DC, Hunt S, Scott CE, Smith J, Vetrie D, Gorman P, Tomlinson IP, Carter NP (2003a) DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. Genes Chromosomes Cancer 36:361-374

Fiegler H, Gribble SM, Burford DC, Carr P, Prigmore E, Porter KM, Clegg S, Crolla JA, Dennis NR, Jacobs P, Carter NP (2003b) Array painting: a method for the rapid analysis of aberrant chromosomes using DNA microarrays. J Med Genet 40:664-670

Fiegler H, Redon R, Carter NP (2007) Construction and use of spotted large-insert clone DNA microarrays for the detection of genomic copy number changes. Nat Protoc 2:577-587

Ford CE, Hamerton JL (1956) The chromosomes of man. Nature 178:1020-1023

Funke B, Edelmann L, McCain N, Pandita RK, Ferreira J, Merscher S, Zohouri M, Cannizzaro L, Shanske A, Morrow BE (1999) Der(22) syndrome and velo-cardio-facial syndrome/DiGeorge syndrome share a 1.5-Mb region of overlap on chromosome 22q11. Am J Hum Genet 64:747-758

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

Giardino D, Finelli P, Gottardi G, Clerici D, Mosca F, Briscioli V, Larizza L (2001) Cryptic subtelomeric translocation t(2;16)(q37;q24) segregating in a family with unexplained stillbirths and a dysmorphic, slightly retarded child. Eur J Hum Genet 9:881-886

Gotter AL, Shaikh TH, Budarf ML, Rhodes CH, Emanuel BS (2004) A palindrome-mediated mechanism distinguishes translocations involving LCR-B of chromosome 22q11.2. Hum Mol Genet 13:103-115

Gray JW, Langlois RG, Carrano AV, Burkhart-Schultz K, Van Dilla MA (1979) High resolution chromosome analysis: one and two parameter flow cytometry. Chromosoma 73:9-27

Gribble SM, Kalaitzopoulos D, Burford DC, Prigmore E, Selzer RR, Ng BL, Matthews NS, Porter KM, Curley R, Lindasy SJ, Baptista J, Richmond TA, Carter NP (2006) Ultra-high resolution array painting facilitates breakpoint sequencing. J Med Genet

Gribble SM, Prigmore E, Burford DC, Porter KM, Ng BL, Douglas EJ, Fiegler H, Carr P, Kalaitzopoulos D, Clegg S, Sandstrom R, Temple IK, Youings SA, Thomas NS, Dennis NR, Jacobs PA, Crolla JA, Carter NP (2005) The complex nature of constitutional de novo apparently balanced translocations in patients presenting with abnormal phenotypes. J Med Genet 42:8-16

Haider S, Matsumoto R, Kurosawa N, Wakui K, Fukushima Y, Isobe M (2006) Molecular characterization of a novel translocation t(5;14)(q21;q32) in a patient with congenital abnormalities. J Hum Genet 51:335-340

Hammond KL, Hanson IM, Brown AG, Lettice LA, Hill RE (1998) Mammalian and Drosophila dachshund genes are related to the Ski proto-oncogene and are expressed in eye and limb. Mech Dev 74:121-131
Harnett GB, Farr TJ, Pietroboni GR, Bucens MR (1990) Frequent shedding of human herpesvirus 6 in saliva. J Med Virol 30:128-130
Hill AS, MacCallum PK, Young BD, Lillington DM (2003) Molecular cloning of a constitutional t(7;22) translocation associated with risk of hematological malignancy. Genes Chromosomes Cancer 38:260-264
Holder JL, Jr., Butte NF, Zinn AR (2000) Profound obesity associated with a balanced translocation that disrupts the SIM1 gene. Hum Mol Genet 9:101-108
Holliday R (1964) A mechanism for gene conversion in fungi. Genet Res 5:282-304
Holmes SE, Riazi MA, Gong W, McDermid HE, Sellinger BT, Hua A, Chen F, Wang Z, Zhang G, Roe B, Gonzalez I, McDonald-McGinn DM, Zackai E, Emanuel BS, Budarf ML (1997) Disruption of the clathrin heavy chain-like gene (CLTCL) associated with features of DGS/VCFS: a balanced (21;22)(p12;q11) translocation. Hum Mol Genet 6:357-367
Hua SB, Qiu M, Chan E, Zhu L, Luo Y (1997) Minimum length of sequence homology required for in vivo cloning by homologous recombination in yeast. Plasmid 38:91-96
Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C (2004) Detection of large-scale variation in the human genome. Nat Genet 36:949-951
IHGSC (2004) Finishing the euchromatic sequence of the human genome. Nature 431:931-945
Ikegawa S, Masuno M, Kumano Y, Okawa A, Isomura M, Koyama K, Okui K, Makita Y, Sasaki M, Kohdera U, Okuda M, Koyama H, Ohashi H, Tajiri H, Imaizumi K, Nakamura Y (1999) Cloning of translocation breakpoints associated with Shwachman syndrome and identification of a candidate gene. Clin Genet 55:466-472
Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nat Genet 6:84-89
Ishikawa_Brush Y, Powell JF, Bolton P, Miller AP, Francis F, Willard HF, Lehrach H, Monaco AP (1997) Autism and multiple exostoses associated with an X;8 translocation occurring within the GRPR gene and 3' to the SDC2 gene. Human Molecular Genetics 6:1241-1250
Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP, Snijders A, Albertson DG, Pinkel D, Marra MA, Ling V, MacAulay C, Lam WL (2004) A tiling resolution DNA microarray with complete coverage of the human genome. Nat Genet 36:299-303
Jacobs PA (1981) Mutation rates of structural chromosome rearrangements in man. Am J Hum Genet 33:44-54

Jacobs PA, Browne C, Gregson N, Joyce C, White H (1992) Estimates of the frequency of chromosome abnormalities detectable in unselected newborns using moderate levels of banding. J Med Genet 29:103-108
Jeffreys AJ (1987) Highly variable minisatellites and DNA fingerprints. Biochem Soc Trans 15:309-317
Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. Nature 314:67-73
Jeffries AR, Mungall AJ, Dawson E, Halls K, Langford CF, Murray RM, Dunham I, Powell JF (2003) beta-1,3-Glucuronyltransferase-1 gene implicated as a candidate for a schizophrenia-like psychosis through molecular analysis of a balanced translocation. Mol Psychiatry 8:654-663
Kalaitzopoulos D (2006) Molecular characterisation and computational analysis of constitutional chromosome rearrangements. PhD Thesis
Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 258:818-821
Kas E, Laemmli UK (1992) In vivo topoisomerase II cleavage of the Drosophila histone and satellite III repeats: DNA sequence and structural characteristics. Embo J 11:705-716
Kehrer-Sawatzki H, Haussler J, Krone W, Bode H, Jenne DE, Mehnert KU, Tummers U, Assum G (1997) The second case of a t(17;22) in a family with neurofibromatosis type 1: sequence analysis of the breakpoint regions. Hum Genet 99:237-247
Kim UJ, Birren BW, Slepak T, Mancino V, Boysen C, Kang HL, Simon MI, Shizuya H (1996) Construction and characterization of a human bacterial artificial chromosome library. Genomics 34:213-218
Kirchhoff M, Gerdes T, Maahr J, Rose H, Bentz M, Dohner H, Lundsteen C (1999) Deletions below 10 megabasepairs are detected in comparative genomic hybridization by standard reference intervals. Genes Chromosomes Cancer 25:410-413
Klar J, Asling B, Carlsson B, Ulvsback M, Dellsen A, Strom C, Rhedin M, Forslund A, Anneren G, Ludvigsson JF, Dahl N (2005) RAR-related orphan receptor A isoform 1 (RORa1) is disrupted by a balanced translocation t(4;15)(q22.3;q21.3) associated with severe obesity. Eur J Hum Genet 13:928-934
Klopocki E, Schulze H, Strauss G, Ott CE, Hall J, Trotier F, Fleischhauer S, Greenhalgh L, Newbury-Ecob RA, Neumann LM, Habenicht R, Konig R, Seemanova E, Megarbane A, Ropers HH, Ullmann R, Horn D, Mundlos S (2007) Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome. Am J Hum Genet 80:232-240
Knowles WA, Gardner SD (1988) High prevalence of antibody to human herpesvirus-6 and seroconversion associated with rash in two infants. Lancet 2:912-913
Kondo K, Kondo T, Okuno T, Takahashi M, Yamanishi K (1991) Latent human herpesvirus 6 infection of human monocytes/macrophages. J Gen Virol 72 (Pt 6):1401-1408

Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders AC, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M (2007) Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome. Science

Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders ACE, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M (In press) Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome. Science

Korenberg JR, Rykowski MC (1988) Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. Cell 53:391-400

Krebs I, Weis I, Hudler M, Rommens JM, Roth H, Scherer SW, Tsui LC, Fuchtbauer EM, Grzeschik KH, Tsuji K, Kunz J (1997) Translocation breakpoint maps 5 kb 3' from TWIST in a patient affected with Saethre-Chotzen syndrome. Hum Mol Genet 6:1079-1086

Krowczynska AM, Rudders RA, Krontiris TG (1990) The human minisatellite consensus at breakpoints of oncogene translocations. Nucleic Acids Res 18:1121-1127

Kumar A, Becker LA, Depinet TW, Haren JM, Kurtz CL, Robin NH, Cassidy SB, Wolff DJ, Schwartz S (1998) Molecular characterization and delineation of subtle deletions in de novo "balanced" chromosomal rearrangements. Hum Genet 103:173-178

Kurahashi H, Emanuel BS (2001) Long AT-rich palindromes and the constitutional t(11;22) breakpoint. Human Molecular Genetics 10:2605-2617

Kurahashi H, Sakamoto M, Ono J, Honda A, Okada S, Nakamura Y (1998) Molecular cloning of the chromosomal breakpoint in the LIS1 gene of a patient with isolated lissencephaly and balanced t(8;17). Human Genetics 103:189-192

Kurotaki N, Stankiewicz P, Wakui K, Niikawa N, Lupski JR (2005) Sotos syndrome common deletion is mediated by directly oriented subunits within inverted Sos-REP low-copy repeats. Hum Mol Genet 14:535-542

Lalli C, Galasso C, Lo Castro A, Nardone AM, Di Paolo A, Curatolo P (2007) Interstitial deletion of a proximal 3p: A clinically recognisable syndrome. Brain Dev 29:312-316

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409:860-921

Landry JR, Mager DL (2002) Widely spaced alternative promoters, conserved between human and rodent, control expression of the Opitz syndrome gene MID1. Genomics 80:499-508

Langlois RG, Yu LC, Gray JW, Carrano AV (1982) Quantitative karyotyping of human chromosomes by dual beam flow cytometry. Proc Natl Acad Sci U S A 79:7876-7880

Levy JB, Canoll PD, Silvennoinen O, Barnea G, Morse B, Honegger AM, Huang JT, Cannizzaro LA, Park SH, Druck T, et al. (1993) The cloning of a receptor-type

protein tyrosine phosphatase expressed in the central nervous system. J Biol Chem 268:10573-10581
Lichter P, Joos S, Bentz M, Lampel S (2000) Comparative genomic hybridization: uses and limitations. Semin Hematol 37:348-357
Lin MS, Comings DE, Alfi OS (1977) Optical Studies of the interaction of 4'-6'-diamidino-2-phenylindole with DNA and metaphase chromosomes. Chromosoma 60:15-25
Lindquester GJ, Inoue N, Allen RD, Castelli JW, Stamey FR, Dambaugh TR, O'Brian JJ, Danovich RM, Frenkel N, Pellett PE (1996) Restriction endonuclease mapping and molecular cloning of the human herpesvirus 6 variant B strain Z29 genome. Arch Virol 141:367-379
Lopes J, Ravise N, Vandenberghe A, Palau F, Ionasescu V, Mayer M, Levy N, Wood N, Tachi N, Bouche P, Latour P, Ruberg M, Brice A, LeGuern E (1998) Fine mapping of de novo CMT1A and HNPP rearrangements within CMT1A-REPs evidences two distinct sex-dependent mechanisms and candidate sequences involved in recombination. Hum Mol Genet 7:141-148
Lopez C, Pellett P, Stewart J, Goldsmith C, Sanderlin K, Black J, Warfield D, Feorino P (1988) Characteristics of human herpesvirus-6. J Infect Dis 157:1271-1273
Lukusa T, Devriendt K, Fryns JP (1999) A 3p deletion syndrome in a child with both del(3)(p25-->pter) and dup(17)(q23-->qter). Ann Genet 42:91-94
Lukusa T, Vermeesch JR, Holvoet M, Fryns JP, Devriendt K (2004) Deletion 2q37.3 and autism: molecular cytogenetic mapping of the candidate region for autistic disorder. Genet Couns 15:293-301
Luppi M, Barozzi P, Morris C, Maiorana A, Garber R, Bonacorsi G, Donelli A, Marasca R, Tabilio A, Torelli G (1999) Human herpesvirus 6 latently infects early bone marrow progenitors in vivo. J Virol 73:754-759
Luppi M, Marasca R, Barozzi P, Ferrari S, Ceccherini-Nelli L, Batoni G, Merelli E, Torelli G (1993) Three cases of human herpesvirus-6 latent infection: integration of viral genome in peripheral blood mononuclear cell DNA. J Med Virol 40:44-52
Madan K, Nieuwint AW, van Bever Y (1997) Recombination in a balanced complex translocation of a mother leading to a balanced reciprocal translocation in the child. Review of 60 cases of balanced complex translocations. Hum Genet 99:806-815
Majewski J, Ott J (2000) GT repeats are associated with recombination on human chromosome 22. Genome Res 10:1108-1114
Mansouri MR, Carlsson B, Davey E, Nordenskjold A, Wester T, Anneren G, Lackgren G, Dahl N (2006) Molecular genetic analysis of a de novo balanced translocation t(6;17)(p21.31;q11.2) associated with hypospadias and anorectal malformation. Hum Genet 119:162-168
Mantripragada KK, Buckley PG, Jarbo C, Menzel U, Dumanski JP (2003) Development of NF2 gene specific, strictly sequence defined diagnostic microarray for deletion detection. J Mol Med 81:443-451
Marui T, Koishi S, Funatogawa I, Yamamoto K, Matsumoto H, Hashimoto O, Nanba E, Kato C, Ishijima M, Watanabe K, Kasai K, Kato N, Sasaki T (2005) No

association of FOXP2 and PTPRZ1 on 7q31 with autism from the Japanese population. Neurosci Res
Matsumoto N, David DE, Johnson EW, Konecki D, Burmester JK, Ledbetter DH, Weber JL (2000) Breakpoint sequences of an 1;8 translocation in a family with Gilles de la Tourette syndrome. Eur J Hum Genet 8:875-883
McMullan TW, Crolla JA, Gregory SG, Carter NP, Cooper RA, Howell GR, Robinson DO (2002) A candidate gene for congenital bilateral isolated ptosis identified by molecular analysis of a de novo balanced translocation. Hum Genet 110:244-250
McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P (2004) The fine-scale structure of recombination rate variation in the human genome. Science 304:581-584
Meselson MS, Radding CM (1975) A general model for genetic recombination. Proc Natl Acad Sci U S A 72:358-361
Meurer S, Pioch S, Wagner K, Muller-Esterl W, Gross S (2004) AGAP1, a novel binding partner of nitric oxide-sensitive guanylyl cyclase. J Biol Chem 279:49346-49354
Millar JK, Wilson-Annan JC, Anderson S, Christie S, Taylor MS, Semple CA, Devon RS, Clair DM, Muir WJ, Blackwood DH, Porteous DJ (2000) Disruption of two novel genes by a translocation co-segregating with schizophrenia. Hum Mol Genet 9:1415-1423
Moncla A, Philip N, Mattei JF (1995) Blepharophimosis-mental retardation syndrome and terminal deletion of chromosome 3p. J Med Genet 32:245-246
Morgan TH (1910) Sex limited inheritance in Drosophila. Science 32:120-122
Morris C, Luppi M, McDonald M, Barozzi P, Torelli G (1999) Fine mapping of an apparently targeted latent human herpesvirus type 6 integration site in chromosome band 17p13.3. J Med Virol 58:69-75
Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, Meyne J, Ratliff RL, Wu JR (1988) A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. Proc Natl Acad Sci U S A 85:6622-6626
Myers S, Bottolo L, Freeman C, McVean G, Donnelly P (2005) A fine-scale map of recombination rates and hotspots across the human genome. Science 310:321-324
Neri G, Reynolds JF, Westphal J, Hinz J, Daniel A (1984) Interstitial deletion of chromosome 3p: report of a patient and delineation of a proximal 3p deletion syndrome. Am J Med Genet 19:189-193
Nickerson E, Greenberg F, Keating MT, McCaskill C, Shaffer LG (1995) Deletions of the elastin gene at 7q11.23 occur in approximately 90% of patients with Williams syndrome. Am J Hum Genet 56:1156-1161
Nie Z, Stanley KT, Stauffer S, Jacques KM, Hirsch DS, Takei J, Randazzo PA (2002) AGAP1, an endosome-associated, phosphoinositide-dependent ADP-ribosylation factor GTPase-activating protein that affects actin cytoskeleton. J Biol Chem 277:48965-48975
Niimura Y, Gojobori T (2002) In silico chromosome staining: reconstruction of Giemsa bands from the whole human genome sequence. Proc Natl Acad Sci U S A 99:797-802

Nimmakayalu MA, Gotter AL, Shaikh TH, Emanuel BS (2003) A novel sequence-based approach to localize translocation breakpoints identifies the molecular basis of a t(4;22). Hum Mol Genet 12:2817-2825

Nishant KT, Kumar C, Rao MR (2006) HUMHOT: a database of human meiotic recombination hot spots. Nucleic Acids Res 34:D25-28

Nothwang HG, Kim HG, Aoki J, Geisterfer M, Kubart S, Wegner RD, van Moers A, Ashworth LK, Haaf T, Bell J, Arai H, Tommerup N, Ropers HH, Wirth J (2001) Functional hemizygosity of PAFAH1B3 due to a PAFAH1B3-CLK2 fusion gene in a female with mental retardation, ataxia and atrophy of the brain. Hum Mol Genet 10:797-806

Okuno T, Takahashi K, Balachandra K, Shiraki K, Yamanishi K, Takahashi M, Baba K (1989) Seroepidemiology of human herpesvirus 6 infection in normal children and adults. J Clin Microbiol 27:651-653

Osborne LR, Li M, Pober B, Chitayat D, Bodurtha J, Mandel A, Costa T, Grebe T, Cox S, Tsui LC, Scherer SW (2001) A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome. Nat Genet 29:321-325

Osoegawa K, Mammoser AG, Wu C, Frengen E, Zeng C, Catanese JJ, de Jong PJ (2001) A bacterial artificial chromosome library for sequencing the complete human genome. Genome Res 11:483-496

Pan TC, Zhang RZ, Pericak-Vance MA, Tandan R, Fries T, Stajich JM, Viles K, Vance JM, Chu ML, Speer MC (1998) Missense mutation in a von Willebrand factor type A domain of the alpha 3(VI) collagen gene (COL6A3) in a family with Bethlem myopathy. Hum Mol Genet 7:807-812

Park SS, Stankiewicz P, Bi W, Shaw C, Lehoczky J, Dewar K, Birren B, Lupski JR (2002) Structure and evolution of the Smith-Magenis syndrome repeat gene clusters, SMS-REPs. Genome Res 12:729-738

Patau K (1960) The identification of individual chromosomes, especially in man. Am J Hum Genet 12:250-276

Patel PI, Roa BB, Welcher AA, Schoener-Scott R, Trask BJ, Pentao L, Snipes GJ, Garcia CA, Francke U, Shooter EM, Lupski JR, Suter U (1992) The gene for the peripheral myelin protein PMP-22 is a candidate for Charcot-Marie-Tooth disease type 1A. Nat Genet 1:159-165

Patsalis PC, Evangelidou P, Charalambous S, Sismani C (2004) Fluorescence in situ hybridization characterization of apparently balanced translocation reveals cryptic complex chromosomal rearrangements with unexpected level of complexity. Eur J Hum Genet 12:647-653

Pellett PE, Lindquester GJ, Feorino P, Lopez C (1990) Genomic heterogeneity of human herpesvirus 6 isolates. Adv Exp Med Biol 278:9-18

Pietroboni GR, Harnett GB, Bucens MR, Honess RW (1988) Isolation of human herpesvirus 6 from saliva. Lancet 1:1059

Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet 23:41-46

Potocki L, Chen KS, Park SS, Osterholm DE, Withers MA, Kimonis V, Summers AM, Meschino WS, Anyane-Yeboa K, Kashork CD, Shaffer LG, Lupski JR (2000) Molecular mechanism for duplication 17p11.2- the homologous recombination reciprocal of the Smith-Magenis microdeletion. Nat Genet 24:84-87

Puissant H, Azoulay M, Serre JL, Piet LL, Junien C (1988) Molecular analysis of a reciprocal translocation t(5;11) (q11;p13) in a WAGR patient. Hum Genet 79:280-282

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, et al. (2006) Global variation in copy number in the human genome. Nature 444:444-454

Reiter LT, Hastings PJ, Nelis E, De Jonghe P, Van Broeckhoven C, Lupski JR (1998) Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients. Am J Hum Genet 62:1023-1033

Reiter LT, Murakami T, Koeuth T, Gibbs RA, Lupski JR (1997) The human COX10 gene is disrupted during homologous recombination between the 24 kb proximal and distal CMT1A-REPs. Hum Mol Genet 6:1595-1603

Richards CS, Watkins SC, Hoffman EP, Schneider NR, Milsark IW, Katz KS, Cook JD, Kunkel LM, Cortada JM (1990) Skewed X inactivation in a female MZ twin results in Duchenne muscular dystrophy. Am J Hum Genet 46:672-681

Roberts T, Chernova O, Cowell JK (1998) NB4S, a member of the TBC1 domain family of genes, is truncated as a result of a constitutional t(1;10)(p22;q21) chromosome translocation in a patient with stage 4S neuroblastoma. Hum Mol Genet 7:1169-1178

Rodriguez-Perales S, Melendez B, Gribble SM, Valle L, Carter NP, Santamaria I, Conde L, Urioste M, Benitez J, Cigudosa JC (2004) Cloning of a new familial t(3;8) translocation associated with conventional renal cell carcinoma reveals a 5 kb microdeletion and no gene involved in the rearrangement. Hum Mol Genet 13:983-990

Rowley JD (1973) Identificaton of a translocation with quinacrine fluorescence in a patient with acute leukemia. Ann Genet 16:109-112

Sakano H, Huppi K, Heinrich G, Tonegawa S (1979) Sequences at the somatic recombination sites of immunoglobulin light-chain genes. Nature 280:288-294

Salahuddin SZ, Ablashi DV, Markham PD, Josephs SF, Sturzenegger S, Kaplan M, Halligan G, Biberfeld P, Wong-Staal F, Kramarsky B, et al. (1986) Isolation of a new virus, HBLV, in patients with lymphoproliferative disorders. Science 234:596-601

Samulski RJ, Zhu X, Xiao X, Brook JD, Housman DE, Epstein N, Hunter LA (1991) Targeted integration of adeno-associated virus (AAV) into human chromosome 19. Embo J 10:3941-3950

Sander M, Hsieh TS (1985) Drosophila topoisomerase II double-strand DNA cleavage: analysis of DNA sequence homology at the cleavage site. Nucleic Acids Res 13:1057-1072

Saxinger C, Polesky H, Eby N, Grufferman S, Murphy R, Tegtmeir G, Parekh V, Memon S, Hung C (1988) Antibody reactivity with HBLV (HHV-6) in U.S. populations. J Virol Methods 21:199-208

Schoumans J, Anderlid BM, Blennow E, Teh BT, Nordenskjold M (2004) The performance of CGH array for the detection of cryptic constitutional chromosome imbalances. J Med Genet 41:198-202

Schule B, Albalwi M, Northrop E, Francis DI, Rowell M, Slater HR, Gardner RJ, Francke U (2005) Molecular breakpoint cloning and gene expression studies of a novel translocation t(4;15)(q27;q11.2) associated with Prader-Willi syndrome. BMC Med Genet 6:18

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004) Large-scale copy number polymorphism in the human genome. Science 305:525-528

Selzer RR, Richmond TA, Pofahl NJ, Green RD, Eis PS, Nair P, Brothman AR, Stallings RL (2005) Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. Genes Chromosomes Cancer 44:305-319

Service RF (2006) Gene sequencing. The race for the $1000 genome. Science 311:1544-1546

Shaffer LG, Lupski JR (2000) Molecular mechanisms for constitutional chromosomal rearrangements in humans. Annu Rev Genet 34:297-329

Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, Stewart H, Price SM, Blair E, Hennekam RC, Fitzpatrick CA, Segraves R, Richmond TA, Guiver C, Albertson DG, Pinkel D, Eis PS, Schwartz S, Knight SJ, Eichler EE (2006) Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nat Genet 38:1038-1042

Shaw CJ, Shaw CA, Yu W, Stankiewicz P, White LD, Beaudet AL, Lupski JR (2004) Comparative genomic hybridisation using a proximal 17p BAC/PAC array detects rearrangements responsible for four genomic disorders. J Med Genet 41:113-119

Shaw-Smith C, Pittman AM, Willatt L, Martin H, Rickman L, Gribble S, Curley R, Cumming S, Dunn C, Kalaitzopoulos D, Porter K, Prigmore E, Krepischi-Santos AC, Varela MC, Koiffmann CP, Lees AJ, Rosenberg C, Firth HV, de Silva R, Carter NP (2006) Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. Nat Genet 38:1032-1037

Shaw-Smith C, Redon R, Rickman L, Rio M, Willatt L, Fiegler H, Firth H, Sanlaville D, Winter R, Colleaux L, Bobrow M, Carter NP (2004) Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. J Med Genet 41:241-248

Shoichet SA, Hoffmann K, Menzel C, Trautmann U, Moser B, Hoeltzenbein M, Echenne B, Partington M, Van Bokhoven H, Moraine C, Fryns JP, Chelly J, Rott HD, Ropers HH, Kalscheuer VM (2003) Mutations in the ZNF41 gene are associated

with cognitive deficits: identification of a new candidate for X-linked mental retardation. Am J Hum Genet 73:1341-1354
Singh GB, Kramer JA, Krawetz SA (1997) Mathematical model to predict regions of chromatin attachment to the nuclear matrix. Nucleic Acids Res 25:1419-1425
Smit AF (1996) The origin of interspersed repeats in the human genome. Curr Opin Genet Dev 6:743-748
Smith AC, McGavran L, Robinson J, Waldstein G, Macfarlane J, Zonona J, Reiss J, Lahr M, Allen L, Magenis E (1986) Interstitial deletion of (17)(p11.2p11.2) in nine patients. Am J Med Genet 24:393-414
Smith BL, Baumgarten R, Nielsen S, Raben D, Zeidel ML, Agre P (1993) Concurrent expression of erythroid and renal aquaporin CHIP and appearance of water channel activity in perinatal rats. J Clin Invest 92:2035-2041
Smith GR, Kunes SM, Schultz DW, Taylor A, Triman KL (1981) Structure of chi hotspots of generalized recombination. Cell 24:429-436
Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, Law S, Myambo K, Palmer J, Ylstra B, Yue JP, Gray JW, Jain AN, Pinkel D, Albertson DG (2001) Assembly of microarrays for genome-wide measurement of DNA copy number. Nat Genet 29:263-264
Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H, Cremer T, Lichter P (1997) Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer 20:399-407
Spitz F, Montavon T, Monso-Hinard C, Morris M, Ventruto ML, Antonarakis S, Ventruto V, Duboule D (2002) A t(2;8) balanced translocation with breakpoints near the human HOXD complex causes mesomelic dysplasia and vertebral defects. Genomics 79:493-498
Spitzner JR, Muller MT (1988) A consensus sequence for cleavage by vertebrate DNA topoisomerase II. Nucleic Acids Res 16:5533-5556
Stankiewicz P, Lupski JR (2002) Genome architecture, rearrangements and genomic disorders. Trends Genet 18:74-82
Strachan T, Read AP (1999) Human Molecular Genetics. John Wiley & Sons, Inc.
Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavare S, Deloukas P, Hurles ME, Dermitzakis ET (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 315:848-853
Sugawara H, Egashira M, Harada N, Jakobs TC, Yoshiura K, Kishino T, Ohta T, Niikawa N, Matsumoto N (2002) Breakpoint analysis of a familial balanced translocation t(2;8)(q31;p21) associated with mesomelic dysplasia. J Med Genet 39:E34
Tagariello A, Heller R, Greven A, Kalscheuer VM, Molter T, Rauch A, Kress W, Winterpacht A (2006) Balanced translocation in a patient with craniosynostosis disrupts the SOX6 gene and an evolutionarily conserved non-transcribed region. J Med Genet 43:534-540
Tedder RS, Briggs M, Cameron CH, Honess R, Robertson D, Whittle H (1987) A novel lymphotropic herpesvirus. Lancet 2:390-392

Thomson BJ, Dewhurst S, Gray D (1994) Structure and heterogeneity of the a sequences of human herpesvirus 6 strain variants U1102 and Z29 and identification of human telomeric repeat sequences at the genomic termini. J Virol 68:3007-3014
Tijo JH, Levan A (1956) The chromosome number of man. Hereditas 42:1-6
Torelli G, Barozzi P, Marasca R, Cocconcelli P, Merelli E, Ceccherini-Nelli L, Ferrari S, Luppi M (1995) Targeted integration of human herpesvirus 6 in the p arm of chromosome 17 of human peripheral blood mononuclear cells in vivo. J Med Virol 46:178-188
Toriello HV, Glover TW, Takahara K, Byers PH, Miller DE, Higgins JV, Greenspan DS (1996) A translocation interrupts the COL5A1 gene in a patient with Ehlers-Danlos syndrome and hypomelanosis of Ito. Nat Genet 13:361-365
Traver CN, Klapholz S, Hyman RW, Davis RW (1989) Rapid screening of a human genomic library in yeast artificial chromosomes for single-copy sequences. Proc Natl Acad Sci U S A 86:5898-5902
Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE (2005) Fine-scale structural variation of the human genome. Nat Genet
van Bakel I, Holt S, Craig I, Boyd Y (1995) Sequence analysis of the breakpoint regions of an X;5 translocation in a female with Duchenne muscular dystrophy. Am J Hum Genet 57:329-336
Velagaleti GV, Bien-Willner GA, Northup JK, Lockhart LH, Hawkins JC, Jalal SM, Withers M, Lupski JR, Stankiewicz P (2005) Position Effects Due to Chromosome Breakpoints that Map ~900 Kb Upstream and ~1.3 Mb Downstream of SOX9 in Two Patients with Campomelic Dysplasia. Am J Hum Genet 76:652-662
Vermeesch JR, Fiegler H, de Leeuw N, Szuhai K, Schoumans J, Ciccone R, Speleman F, Rauch A, Clayton-Smith J, Van Ravenswaaij C, Sanlaville D, Patsalis PC, Firth H, Devriendt K, Zuffardi O (2007) Guidelines for molecular karyotyping in constitutional genetic diagnosis. Eur J Hum Genet
Vervoort VS, Viljoen D, Smart R, Suthers G, DuPont BR, Abbott A, Schwartz CE (2002) Sorting nexin 3 (SNX3) is disrupted in a patient with a translocation t(6;13)(q21;q12) and microcephaly, microphthalmia, ectrodactyly, prognathism (MMEP) phenotype. J Med Genet 39:893-899
Visser R, Shimokawa O, Harada N, Kinoshita A, Ohta T, Niikawa N, Matsumoto N (2005) Identification of a 3.0-kb major recombination hotspot in patients with sotos syndrome who carry a common 1.9-Mb microdeletion. Am J Hum Genet 76:52-67
Vissers LE, de Vries BB, Osoegawa K, Janssen IM, Feuth T, Choy CO, Straatman H, van der Vliet W, Huys EH, van Rijk A, Smeets D, van Ravenswaaij-Arts CM, Knoers NV, van der Burgt I, de Jong PJ, Brunner HG, van Kessel AG, Schoenmakers EF, Veltman JA (2003) Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. Am J Hum Genet 73:1261-1270

Volik S, Zhao S, Chin K, Brebner JH, Herndon DR, Tao Q, Kowbel D, Huang G, Lapuk A, Kuo WL, Magrane G, De Jong P, Gray JW, Collins C (2003) End-sequence profiling: sequence-based analysis of aberrant genomes. Proc Natl Acad Sci U S A 100:7696-7701

Wahlstrom J, Uller A, Johannesson T, Holmqvist D, Darnfors C, Vujic M, Tonnby B, Hagberg B, Martinsson T (1999) Congenital variant Rett syndrome in a girl with terminal deletion of chromosome 3p. J Med Genet 36:343-345

Waldman AS, Liskay RM (1988) Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. Mol Cell Biol 8:5350-5357

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 49:995-1013

Ward KN, Gray JJ, Efstathiou S (1989) Brief report: primary human herpesvirus 6 infection in a patient following liver transplantation from a seropositive donor. J Med Virol 28:69-72

Wassink TH, Piven J, Vieland VJ, Jenkins L, Frantz R, Bartlett CW, Goedken R, Childress D, Spence MA, Smith M, Sheffield VC (2005) Evaluation of the chromosome 2q37.3 gene CENTG2 as an autism susceptibility gene. Am J Med Genet B Neuropsychiatr Genet 136:36-44

Watt VM, Ingles CJ, Urdea MS, Rutter WJ (1985) Homology requirements for recombination in Escherichia coli. Proc Natl Acad Sci U S A 82:4768-4772

Weksberg R, Shuman C, Caluseriu O, Smith AC, Fei YL, Nishikawa J, Stockley TL, Best L, Chitayat D, Olney A, Ives E, Schneider A, Bestor TH, Li M, Sadowski P, Squire J (2002) Discordant KCNQ1OT1 imprinting in sets of monozygotic twins discordant for Beckwith-Wiedemann syndrome. Hum Mol Genet 11:1317-1325

Willett-Brozick JE, Savul SA, Richey LE, Baysal BE (2001) Germ line insertion of mtDNA at the breakpoint junction of a reciprocal constitutional translocation. Hum Genet 109:216-223

Wood S, Schertzer M, Drabkin H, Patterson D, Longmire JL, Deaven LL (1992) Characterization of a human chromosome 8 cosmid library constructed from flow-sorted chromosomes. Cytogenet Cell Genet 59:243-247

Wu K, Yang Y, Wang C, Davoli MA, D'Amico M, Li A, Cveklova K, Kozmik Z, Lisanti MP, Russell RG, Cvekl A, Pestell RG (2003) DACH1 inhibits transforming growth factor-beta signaling through binding Smad4. J Biol Chem 278:51673-51684

Wyman AR, White R (1980) A highly polymorphic locus in human DNA. Proc Natl Acad Sci U S A 77:6754-6758

Yoshiura K, Machida J, Daack-Hirsch S, Patil SR, Ashworth LK, Hecht JT, Murray JC (1998) Characterization of a novel gene disrupted by a balanced chromosomal translocation t(2;19)(q11.2;q13.3) in a family with cleft lip and palate. Genomics 54:231-240